

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:42:58 ; Search time 69 Seconds  
(without alignments)  
6850.753 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VIKELKLFRRITMSDNEF.....PRKILGRTEKTKPKTERP 1573

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8879	100.0	1673	3	AAY95707 Cosmid cH
2	1638	18.4	1584	2	AAY33727 Photorhab
3	1583	17.8	1787	6	ABM67433 Photorhab
4	1569	17.7	1590	6	ABM67283 Photorhab
5	432	4.9	2334	5	ABG31849 Human kin
6	386	4.3	2234	6	ABU18641 Protein e
7	365	4.1	1400	7	ADCO1365 Enterohae
8	353.5	4.0	843	6	ABU48290 Protein e
9	350	3.9	1515	6	ABU19676 Protein e
10	342	3.9	1397	6	ABU15135 Protein e
11	341.5	3.8	1404	7	ADCO0960 Enterohae
12	341	3.8	1394	7	ADCO1426 Enterohae
13	340.5	3.8	1411	6	ABU15136 Protein e
14	339	3.8	1426	3	ABU15983 E. coli p
15	339	3.8	1426	6	ABU14693 Protein e
16	331.5	3.7	1377	4	AAU34791 E. coli c
17	331.5	3.7	1377	4	AAU34791 E. coli g
18	331.5	3.7	1377	6	ABU14807 Protein e
19	328	3.7	1572	6	ABU41491 Protein e
20	324.5	3.7	1429	6	ABM69806 Photorhab
21	322	3.6	1385	6	ABU40318 Protein e
22	317.5	3.6	1584	6	ABM66973 Photorhab
23	317.5	3.6	1585	6	ABU16634 Protein e
24	314	3.5	1395	6	ABU22662 Protein e
25	313	3.5	1627	6	ADA35317 Acinetoba

26 307.5 3.5 932 6 ABU16693 Protein e  
27 306.5 3.5 1530 6 ABU40251 Protein e  
28 298.5 3.4 1531 6 ABU21384 Protein e  
29 298 3.4 885 4 AAU52772 Propionib  
30 298 3.4 885 6 ABM49291 Propionib  
31 298 3.4 920 6 ABM65127 Propionib  
32 297 3.3 1512 6 ABU50462 Protein e  
33 291.5 3.3 1253 6 ABU40990 Protein e  
34 289.5 3.3 1364 6 ABU47385 Protein e  
35 285.5 3.2 1596 6 ADA34255 Acinetoba  
36 285 3.2 2613 5 ABP53589 Human NOV  
37 285 3.2 2628 5 ABP53588 Human NOV  
38 285 3.2 2721 5 ABP53587 Human NOV  
39 285 3.2 2725 5 ABP53586 Human NOV  
40 280 3.2 1317 4 AAU33622 Pseudomon  
41 280 3.2 1317 6 ABU15571 Protein e  
42 279.5 3.1 944 6 ABU19882 Protein e  
43 278 3.1 985 6 ABU15137 Protein e  
44 276.5 3.1 1438 6 ABU50436 Protein e  
45 273.5 3.1 1688 6 ABR58344 XM\_047995

## ALIGNMENTS

RESULT 1  
AAY95707  
ID AAY95707 standard; protein; 1673 AA.

XX AAY95707;

DT 25-OCT-2000 (first entry)

XX Cosmid CHRM5 encoded protein Pl4-2f.

XX Cosmid CHRM5; nematocite; nematode; biological control agent;  
XX transgenic plant; helminthiasis; Pl4-2f.

XX Xenorhabdus bovienii.

XX WC2000042855-A1.

XX 27-JUL-2000.

XX 24-JAN-2000; 2000WO-GB000219.

XX 22-JAN-1999; 99GB-00001499.

XX (HORT-) HORTICULTURE RES INT.

XX Morgan JAW, Jarrett P, Ellis D, Ousley MA;

XX WPI; 2000-499157/44.

XX N-PSDB; AAA50029.

XX Novel composition used to control parasitic nematodes, especially in a  
XX plants such as maize, cotton, soya, and rice, comprises a bacterium which  
XX is a symbiont of an entomopathogenic nematode.

XX Example 6; Page 42-43; 74pp; English.

XX The present sequence is that of protein Pl4-2f encoded by an open reading  
XX frame identified in cosmid CHRM5 (see AAA50029). CHRM5 was obtained by  
XX ligating Xenorhabdus bovienii strain 173 (NCIMB 40986) Sau3A-digested DNA  
XX fragments into the BamHI site of the Stratagene cosmid vector Supercoi,  
XX packaging into Escherichia coli XL Blue 1, and screening for nematocidal  
XX activity against Caenorhabditis elegans. Analysis of the DNA indicated a  
XX number of open reading frames for which the corresponding protein  
XX sequences were determined (see AAY95685-Y95735). Nematodes can be  
XX controlled through the use of bacteria associated symbiotically with an  
XX entomopathogenic nematode. Such bacteria include Xenorhabdus and  
XX Photorhabdus spp. such as X. bovienii strain 173. The symbiont bacteria,  
XX an engineered bacterium, or a nematocidal protein obtained from such

CC bacteria, particularly P13-1f (see AAY95706) or P14-2f can be used to  
 CC control helminthiasis in a human or domesticated animal or for the  
 CC control of plant pathogen nematodes. Also claimed are vectors for  
 CC expressing nematocidal proteins in host cells, and transgenic plants  
 XX  
 SQ

Sequence 1673 AA;

Query Match	100.0%;	Score 8879;	DB 3;	Length 1673;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1673;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	VYIKFLKLPRIITMSDNNNEFFTOANNFTSAVSGGVDPRGLYNIQITIGHIVGNGNLGPT	60	
DB	1	VYIKFLKLPRIITMSDNNNEFFTOANNFTSAVSGGVDPRGLYNIQITIGHIVGNGNLGPT	60	
QY	61	LPTLSYSPLNKTDIGFIGNFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKKLD	120	
DB	61	LPTLSYSPLNKTDIGFIGNFGLSVYDRKNSLSLSTGENYKVIETDKTVLQOKKLD	120	
QY	121	NLRFKDLKENCYRIITHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDMFEATOP	180	
DB	121	NLRFKDLKENCYRIITHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDMFEATOP	180	
QY	181	RLNRIYDDLGDHDIPLNLEYQSLIKITLTPFGQKEGYRTELRLFNQLNSIHNFSLGN	240	
DB	181	RLNRIYDDLGDHDIPLNLEYQSLIKITLTPFGQKEGYRTELRLFNQLNSIHNFSLGN	240	
QY	241	ENPLTWSFGYTPIGKXGILGOWITMTAPGGLKETVYNSNNQGHFFQSANLPVLPVYT	300	
DB	241	ENPLTWSFGYTPIGKXGILGOWITMTAPGGLKETVYNSNNQGHFFQSANLPVLPVYT	300	
QY	301	LMQVPGAGOPAIQAEYSYTSNHYVGGSGNGIWNKLNLYGLMTEYNYGTESSRYKOK	360	
DB	301	LMQVPGAGOPAIQAEYSYTSNHYVGGSGNGIWNKLNLYGLMTEYNYGTESSRYKOK	360	
QY	361	EGHDQIVRIERTYNNYHLLTSECKQNGVIOITETAYYAIIGHNFDSPQFOLPKTKTE	420	
DB	361	EGHDQIVRIERTYNNYHLLTSECKQNGVIOITETAYYAIIGHNFDSPQFOLPKTKTE	420	
QY	421	TWRSADNSVSEITETTFDESNGNPLTKVTKDKKTKQIISPSHWHYPPAGEVNCNCPPEP	480	
DB	421	TWRSADNSVSEITETTFDESNGNPLTKVTKDKKTKQIISPSHWHYPPAGEVNCNCPPEP	480	
QY	481	YGTRFVKKIIQTPYDSEFKDDPEKFIQVRYSLIGSQSHVTLKIBERHYSATQLNSTLF	540	
DB	481	YGTRFVKKIIQTPYDSEFKDDPEKFIQVRYSLIGSQSHVTLKIBERHYSATQLNSTLF	540	
QY	541	QYNTDSELGRLLKQTECTKGNGKTYSVVHFTYTKQDDTLQOQSHSITTHDNFTIHSQ	600	
DB	541	QYNTDSELGRLLKQTECTKGNGKTYSVVHFTYTKQDDTLQOQSHSITTHDNFTIHSQ	600	
QY	601	VRGRYTRGLFSDTDDKDIOTQMSYDKLGRLLTSLNSGTPYANTUTYVELNNIQQDNRP	660	
DB	601	VRGRYTRGLFSDTDDKDIOTQMSYDKLGRLLTSLNSGTPYANTUTYVELNNIQQDNRP	660	
QY	661	PFVITTTDNGNOLRNEFDGAGRHSVQCLKDSGDKGFYTHIQOYDEQGRHHTSYSDY	720	
DB	661	PFVITTTDNGNOLRNEFDGAGRHSVQCLKDSGDKGFYTHIQOYDEQGRHHTSYSDY	720	
QY	721	LITNGRQOTDPDKVHLMSKSYDNWQGIANTHWSYGVSEKIIVDPITLTLATKQLOSNVNV	780	
DB	721	LITNGRQOTDPDKVHLMSKSYDNWQGIANTHWSYGVSEKIIVDPITLTLATKQLOSNVNV	780	
QY	781	QTKGEVITYTPSQOPIQITLPEAGHLQSCHTLTRDGDWRVKEFDAIGCQTIYQYDNYN	840	
DB	781	QTKGEVITYTPSQOPIQITLPEAGHLQSCHTLTRDGDWRVKEFDAIGCQTIYQYDNYN	840	
QY	841	RVLIQITLPDGTIVNRKYAFSDTLTIDIRVNGISIGQOTFDGLSRLTQSGDQGRVWAYT	900	
DB	841	RVLIQITLPDGTIVNRKYAFSDTLTIDIRVNGISIGQOTFDGLSRLTQSGDQGRVWAYT	900	
QY	901	YSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTGALLKAVAG	960	

DB	901	YSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTGALLKAVAG	960	
QY	961	QSLTPYYPGSRKLMENINDMKMSYLTWLRGLNGYTDLTGTTIKISRTHGRVTOIKD	1020	
DB	961	QSLTPYYPGSRKLMENINDMKMSYLTWLRGLNGYTDLTGTTIKISRTHGRVTOIKD	1020	
QY	1021	SSIKITLNVDDLNRIHIGSQVTDLATGMLTTTVFDPGLNRIEGRKLCSSGHTLIDQOSW	1080	
DB	1021	SSIKITLNVDDLNRIHIGSQVTDLATGMLTTTVFDPGLNRIEGRKLCSSGHTLIDQOSW	1080	
QY	1081	LKTQOLANRIVKLNGVLOQTEQYSYDSNRNLNQVKCAGCEPTDKYGHISIVTQNTFTYDIY	1140	
DB	1081	LKTQOLANRIVKLNGVLOQTEQYSYDSNRNLNQVKCAGCEPTDKYGHISIVTQNTFTYDIY	1140	
QY	1141	GNITACHITFADGTEDHATFXFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINITD	1200	
DB	1141	GNITACHITFADGTEDHATFXFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINITD	1200	
QY	1201	NHNGTENFTYDGLRLQNGQSVYGYDPLNRLVSKQTDLDCELYYRETMVNEVRNGEM	1260	
DB	1201	NHNGTENFTYDGLRLQNGQSVYGYDPLNRLVSKQTDLDCELYYRETMVNEVRNGEM	1260	
QY	1261	IRLLTGETIIAQORASKVLLTGTDSQOSVILTSDKQNLQVSEAYSAIKGKHSANDASIL	1320	
DB	1261	IRLLTGETIIAQORASKVLLTGTDSQOSVILTSDKQNLQVSEAYSAIKGKHSANDASIL	1320	
QY	1321	GYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLPFGAGGINPYSCIGDPINRSDP	1380	
DB	1321	GYNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLPFGAGGINPYSCIGDPINRSDP	1380	
QY	1381	SGHLSQAWTIGMGIAGLLTIATGMAIAAAGIAAAASTTTALAFALSVTSIT	1440	
DB	1381	SGHLSQAWTIGMGIAGLLTIATGMAIAAAGIAAAASTTTALAFALSVTSIT	1440	
QY	1441	SIVSGALEDASPASSIILGWVSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKST	1500	
DB	1441	SIVSGALEDASPASSIILGWVSMGMAAGLAESAIGKGTKLATHLGAFAEDGENALLKST	1500	
QY	1501	SSSSRIKGVTRSLDREIVRNEEQVTKDHSRGVTDNFMKGEOAILVHGDKGFLYHTE	1560	
DB	1501	SSSSRIKGVTRSLDREIVRNEEQVTKDHSRGVTDNFMKGEOAILVHGDKGFLYHTE	1560	
QY	1561	GNKNGKGPYTRHTPEQLVDYLKDNINVDLTQGGDKPVLHLLSCYKSSGAADKQAKVINR	1620	
DB	1561	GNKNGKGPYTRHTPEQLVDYLKDNINVDLTQGGDKPVLHLLSCYKSSGAADKQAKVINR	1620	
QY	1621	PVIAYSNKPTISQGLARIEKDFLKSTYHSYDPRKILGRTEKTVKPKTRP	1673	
DB	1621	PVIAYSNKPTISQGLARIEKDFLKSTYHSYDPRKILGRTEKTVKPKTRP	1673	

RESULT 2

AAY33727

ID AAY33727 standard; protein; 1584 AA.

XX AAY33727;

XX AC AAY33727;

XX XX 09-NOV-1999 (first entry)

XX DT Photerhabdus luminescens 176 kD insecticidal toxin.

XX DE Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX KW Photerhabdus luminescens.

XX OS Photerhabdus luminescens.

XX XX WO9942589-A2.

XX XX 26-AUG-1999.

XX XX 18-FEB-1999; 99WO-EP001015.

XX XX 20-FEB-1998; 98US-00027080.

XX XX 20-JAN-1999; 99US-0116439F.

PR





Db 1132 TINKITSLDLDGRETETKIRRHNGLKQNKGITDIHISQTFNEQDKITNKL-LHGT 1190  
Qy 1097 LQ-RTEQSYSDRNRLNQKCDGAECPDCKYGHSLVQ--NFTYDIYGNITACHTTTADG 1153  
Db 1191 KOISKEVYTHKRGLETYT-----MELGBERITQCSYKYDHLGNITQ-HSITTEG 1243  
Qy 1154 TEDHATKFNAPT--DPCQLTEVHH-THPDMPDNIRLKYKAGRVINITD-NHGNTENFT 1209  
Db 1244 XTITSTYFGNIGDPCQLIDVSTCTGNSSSLRFTYNGRGALVCENDENNTKIRTWT 1303  
Qy 1210 YDTLRLQNGGCV-----YCYDPLRLV--SQKDET-LDCELYRETMVYN----- 1253  
Db 1304 YDSLGRDITVDALFKVETRYLFDATRLIIRKSEKNGTPVHHDLSCYNSLVHNDVYFG 1363  
Qy 1254 EVRNGEMIR-----LLRGTETIIAQ--QRASKVLLITGDSQCVILTSKQNLQSOE 1302  
Db 1364 EKRNAADRKYNKVGICLGFQSFCHQTPTSVASRYTETATDCKGVIATFQGEDVQHI 1423  
Qy 1303 AYSAYGKHKSTANDASILG-----YNGERADPVSGVTHLNGYRSYDPTLM 1348  
Db 1424 AYSFPGW---VTTEQAMVTAGQQPPHNTAIEBEPFNGEQMDTASAYLLNGYRAYRPDLM 1480  
Qy 1349 RFHPDLSLSPGAGINPVSICLDPINRSPSGHLSWQAMTGHGMGIAGLLITATGCM 1408  
Db 1481 RFTAPDSWSPGAGINAYACGGDPVNLDPDSGHISGHWANITGGIGLLAPFTYGG 1540  
Qy 1409 AIAAGGIAAIASTTTALAFGALSVTSDITSIVSGALEDAPKASSILGWBMGMGAA 1468  
Db 1541 SLELGLGVNAARGLT-----ALDAASGVTAASGALENKPNPETSRRLLGWMBSLGLG-- 1590  
Qy 1469 GLAESAIKGYKLA-----THLGA-----AEDGENALLKKTSE--- 1502  
Db 1591 --LPSMWITGGVSLAOWNVRLTNSRTPYHYPTSLGEVNLGRKSDWVAERSLNSGENW 1648  
Qy 1503 -----SSRIKWG---VTRSLDREIVRNEEQVINDHSRGYTDNFMKGEOAI-----L 1547  
Db 1649 HSEVLNGRTIWGSDTKRGLD---IKYPLEQISERPNSG--DIVLLSGSHGVQNGDWL 1703  
Qy 1548 VHGDKDGFLYH-----TEGNKHNGKGYTRHTPEQLVDYLDKNNIVDLTQGGKPV 1598  
Db 1704 INGSRRGSLHFPFPKSDMTYVGGSWKGRTHVRN-----LATWSEIDFGTLNNGNSHI 1758  
Qy 1599 HLLSCYKGSSEA 1610  
Db 1759 ILGCYCYGRNDQA 1770

RESULT 4  
ABM67283  
ID ABM67283 standard; protein; 1590 AA.  
XX ABM67283;  
AC ABM67283;  
DT 20-NOV-2003 (first entry)  
XX Photorhabdus luminescens protein sequence #380.  
DE  
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KW detection; food; gene expression; plant; animal; microorganism; toxin;  
KW antibiotic; biopesticide; virulence factor; disease model; plague;  
KW whooping cough.  
XX Photorhabdus luminescens.  
OS  
XX WO200294867-A2.  
PN  
XX 28-NOV-2002.  
PD  
XX 07-FEB-2002; 2002WO-IB003040.  
PF  
XX 07-FEB-2001; 2001PR-00001659.  
PR  
XX

PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
PI Buchrieser C;  
XX WPI; 2003-148459/14.  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX Claim 2; SEQ ID NO 380; 1205pp; French.  
XX The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX  
SQ Sequence 1590 AA;  
Query Match 17.7%; Score 1569; DB 6; Length 1590;  
Best Local Similarity 30.8%; Pred. No. 1.9e-100;  
Matches 477; Conservative 236; Mismatches 636; Indels 200; Gaps 57;  
Qy 63 LTUSYGPLNKTDIGFQIGFNGFGLSVVDRKNSLLSLSTGENYK---VIETDKTVLKQOKKL 119  
Db 13 LNLFSFPLATLNNFGFGIGWRFSLTMLDVKTLTFRSNGEQFKCKPLPPNNNDISFCKKL 72  
Qy 120 DNLREKDKLNCYRIIHKSGDIEVLTFGNNAFDLKVPKLLNPAGHAIYIDWNEATQ 179  
Db 73 KDLRVYK-LDSNIFYVYNKGLIETLKRIGSS--DIAKTVALEFPDG----- 116  
Qy 180 PRLNRIYDDLGDHDIPLNLLEYQGLIKTILTL-FPGQKEGYTELRFL-NRQLNSIHNS 237  
Db 117 ---EVFDLTYNRFRFALSEIKYRMTGKTYLKLNSGNN---CTSVEYPPDNNISAKIAFD 169  
Qy 238 LGNENPLTWSFGYTPPIGK-----NGILGQWITSMTAPGLKETVNSNNNG 284  
Db 170 YRNDYDITVTVPYDASGPDSDARFKMTYQTLKQIF-FVISAFRTPTGYVELSVYKEN--G 226  
Qy 285 HHFPQSANLPVLPVYVTLMKQVPGAGQPAIQAEYSYTS-HNYVGGSGNGI-WNNKLDNLYG 342  
Db 227 HKYTDTEISFYAALATIQ---PCNGQPAISKSYEYSSVHNFGLYSSGRTSFDSSQDNLYL 283  
Qy 343 LMTYNYGSTESRRYKDKGHDQIVRIERTYNNHLLTSECKQONGYIQTETAYVAILG 402  
Db 284 VTGKYTYSSIE----RYLNGQNVISVTERVDFKHLMTKEAKTDQNKRIITEITYNEDPS 339  
Qy 403 HNFDSQSPQFLPK-----TKTETWRSADNSYRSEITETTFDESQNPITKVIKDKTKQ 456  
Db 340 KSPSEQENLQPSHVLTRYDLOT-----NTSRESVNIKSDDWGNTLL-ITETSGIOK 393  
Qy 457 IISPSHTWEYPPPAGEVDNCPPEPYGTRFVKKIIQTPYDSEFPDKPEKFIQRYSLIGS 516  
Db 394 -----EYVYVYVNGEGNCPADFLGFSRFLKSVTKQSGPDAAQSVANRVTSYQKPLT 447

QY 517 QSHVTLKIBERHYSATQLNSTL---FOY---NTDKSELGRLLKOTECTKGNGKTSYV 570  
Db 448 FTGAYVK---EYVSKASETIDSKIVRTFNVTNSPTKSH---GSLAKITSWNNO-----QTV 499  
QY 571 HKFTYTKQDDTLQOQSHSIITHDNFTIHRSOVRSRYTGRFLFSDTDKDIOTQMSYDKLGLRL 630  
Db 500 TTFKYEYSDSEMTNSTVTGFGDTHMESKNVTSIVTHRQLRKVDVNVHVTIDQSYDLSGRI 559  
QY 631 LFTLNSGTFPYANTLYDYELANLQDDNRPPEVITTTVANGNQLNEPDGAGRHHVQCCLK 690  
Db 560 IGIIDPGTTHKEIKSIYIYQPCGENDFWP---VMEIDSQGIARRKTHDGMGRICISIEEQ 618  
QY 691 DSDG-----DGFTYTHITQYDEQGRHHTSTYSVLTN-----GRQOTDPPDKVHLMS 737  
Db 619 DDGVMGTSGIYQGYRKLARQYDVLGQVKEISNDMLWDLNPLRLTLP---LVTTK 676  
QY 738 SKSYDNWGOIANTHSYGVSEKLTVDPTILPATK-----QLQSNNVQTGKEVTTYP 791  
Db 677 TYQYDGMGNRYSTESDGRIELEIHPITRTITQGVKGLGMLNIQQNNF-----725  
QY 792 SQPIQITLFDGAGHLSCHTLTRDGDWRVRETDAIGQCTIYQYDYNVRVITQLPDCGT 851  
Db 726 -EQPASIKVYPDGAISYRTYRYDGFCTVTETDAEGYATQIEYDLDFRIVKXLTLPRT 784  
QY 852 IYVRKYAPSTDTLITDIRVNGISIGQOTFDGLSRLTOSQDGRVWAVTYAGNDQCPST 911  
Db 785 ILESAYASFSEELISALNVNGTQGLSLYDGLGRVTRDTVGRKTEYLYGSQGDK-PIQ 843  
QY 912 VITP--DGCFIHYQPELDDAVLQVANSNEITQFSSYNPVTGALLKAVAEGOSLTPI--YY 968  
Db 844 SVTPAHKKNIDLY--ALGVSMTFTTETSQNSFYQKGTGALLSA--TEGVSGNSYF 900  
QY 969 PSGLKLMENIN--DMKMS---YLWTLRGLNGYDITLGTIIOKISRDTHGRVTQIKDSSIK 1024  
Db 901 PSGLVCHESFSDRNKPISSGDYRYTMSGLIQSHKDSFAHDHVSFYDAEGLRVKTEQSSQY 960  
QY 1025 TTLNVDLNRHGSQVDTLATCHMLTTVEPGLNREIGRKLCDSSGHTLIDQOSWLTKQ 1084  
Db 961 ATEYDYNVGRLLTTTTTKDTTSLQATKIEYDFVDFREIKRSLISDFSQV--ITUSYTKN 1019  
QY 1085 QLANRIVKLVGLQRTQYSDSRNLNQKDGACEPTDKYGSIVTQNFYDIYGNIT 1144  
Db 1020 QISQRITSDGVVMKNERYOYDSNQRLSQYQCEGQSVDHTGRVLSQIIVHYDQWGNIK 1079  
QY 1145 ACHTFADCTEDHATFKANPTDPCQTEVHTHDPMDPNRLKYDKAGRVINIT--DWHG 1203  
Db 1080 RLNDTYRDKET--VDYHFSQ--ADPTQLIRI-----TSDKQQLSELYDANG---NLTRDEK 1130  
QY 1204 NTENFTYDLGLR---QNGQGSV---YGYDPLNRLVSQ--KTDTLDCELYRYETMLVNEV 1255  
Db 1131 QT--LIYDONNELVQVQSKGNLVCOQYDQALNKLTQAQVLANGTVNRQ--YVASGNVANVQ 1187  
QY 1256 RNMGMIRLL-----RTGTIIAQORASKVLTGDSQOSVILTSKQNLQOEA 1303  
Db 1188 LGDETITLMSDKQRLGHQSTKNGESVYIYQ-----GTDHNSVTIASQNELMALS 1239  
QY 1304 YSAYGKHSTANDASILGVNGERADPVSGVTHLNGYRSYDPTLMRPHTPDLSFPFGAGG 1363  
Db 1240 YTPYGRSLI---SSLPLNGAQVDPTVGTWFLNGYRVFVFLMRPHFHSPPSWPFGRG 1296  
QY 1364 INPYSYCLGDPINRSDPGHLSWQAWTIGIGIAGLLTIAT--GWAIA-----AAGGIA 1417  
Db 1297 VNPYTYCQDGPINRIDLNGHLSAGGILGIVLGAIGIIVGVVSLGAGAISAAGLAAAGAL 1356  
QY 1418 AAIASSTTALAFAGALSVTSDITTSVSCALDASPKASSILGWVSMGMAAGLAESAII-- 1475  
Db 1357 GAIASSTAFATVATVIGLAASIGIASAALSEKDPKTAGILNWIISTGLCVLSPGISAITP 1416  
QY 1476 -----KGGTKLATH--LQAF--AEDGENALLKSTSESSRIKGVTR 1513  
Db 1417 TSSLIKSARSGSVASTSVIGSVPIEFGEIA-----SRSSR--RWDIALS 1459

RESULT 5  
ABG31849  
ID ABG31849 standard; protein; 2334 AA.  
XX  
AC ABG31849;  
DT 05-NOV-2002 (first entry)  
XX  
DE Human kinase, MEK1.  
XX  
KW Extracellular signal regulated kinase; hyperalgesia; surgery;  
KW Opioid withdrawal; pain sensitisation; analgesic; chronic pain; BRK;  
KW MEK1; human; enzyme.  
XX  
OS Homo sapiens.  
XX  
EN W0200258687-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 25-JAN-2002; 2002WO-US002128.  
XX  
PR 25-JAN-2001; 2001US-0264336P.  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
XX  
PI Gutstein HB;  
XX  
XX WPI; 2002-608414/65.  
XX N-PSDB; ABK30804.  
XX  
XX Reducing or reversing tolerance, physical dependence, hyperalgesia,  
XX withdrawal symptoms, or pain sensitization in patients on analgesics for  
XX chronic pain, comprises inhibition of the extracellular signal-regulated  
XX kinase (ERK).  
XX  
XX Disclosure; Page 156-161; 163pp; English.  
XX  
XX The invention relates to a method of reducing or reversing tolerance,  
XX reducing the risk of physical dependence or hyperalgesia, reducing the  
XX symptoms of opioid withdrawal or inhibiting pain sensitisation in a  
XX patient taking analgesics. The method comprises administering an  
XX analgesic and an extracellular signal-regulated kinase (ERK) inhibitor  
XX comprised in a formulation to reduce or reverse tolerance, risk of  
XX physical dependence, hyperalgesia, symptoms of opioid withdrawal, or  
XX inhibiting pain sensitisation in patients taking analgesics for chronic  
XX pain or those undergoing surgery. The present sequence represents the  
XX amino acid sequence of human MEK1 (not defined)  
XX  
SQ Sequence 2334 AA;  
Query Match 4.9%; Score 432; DB 5; Length 2334;  
Best Local Similarity 20.4%; Pred. No. 3.3e-20;  
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;  
QY 44 IQITLGHIVG--NGNL-----GPTLPLTSLYSPLNKTDIGFGIFNGLSVYDR 90  
Db 896 IDPSSQLNGATGATVIVNEEDLSIDGRGPGIGLSTYNSLDSSDHLFGQW----YADA 950  
QY 91 KNSLLSLSTGENYKVIETDKTVLQOKKLDN-----LRFKDLKENCYRIHKSGDIEVL 145  
Db 951 ETSVISTDQGAMY--IDEDATTHRTFKADGTGYQPTGVYLELTETADQFILKTKD---- 1004  
QY 146 TGFNNNAFLKVPKLL-----NPAGHAIYIDWN-----FEATQPLNRIYDDLDGH-- 192  
Db 1005 ---QTWAYNKKGKLVQVGDHNNATVYVYNDKNQLTAITDASGRKLTFTYDE--NGHVT 1060  
QY 193 -----DIPLLNLEY--QGLI-----KTILTLPFG 214  
Db 1061 SITGPKNKVTSYENDLLKKVYTDGTVDYDSYDSEGLRVKQYGSANSTKAPVFTEY-- 1118  
QY 215 QXEGYRTELRLNRLNLSIHNF-----LCNENPLTWSFGYTPIGKNGILGQWITSMT 267

Db 1119 QYSHRLEKAINAKKETYVSYDADKTKLLTLPQGRKVOYGYNEAGNP-----IQVID 1172  
QY 268 APGLKETVN--YNNN-----QGHFPQSANLPVLPVYTLMKQVPGAG 309  
Db 1173 DAEGLKITNTKYEGNWNVEDVPNDVGTGKATESYQYDKGN-----VTSVKDAYGT- 1225  
QY 310 QPALQABYSYSHNYVGGGNGIWNKLDNLYGLMTEYNYGSTESRRYKDKGHDOIVRI 369  
Db 1226 -----ETYEYNNKNDV-----TKQKOTEGNVTDIADGLDASETDQSGKSSAAV 1271  
QY 370 ERTYNNVHLLTSECKQONGVIQITETAYYAIIGHNFDSPQSFQPLPKTKTETWRSADNSY 429  
Db 1272 YDKYGNQIQSSKDLASTNLIK-----DGSPEAKSGWNLTASKD----- 1311  
QY 430 RSEITETTFDESIGNPLTKVTKDKTKIISPS-----THWEYPPAGEVDNCPPEPYGFTR 485  
Db 1312 RRTIS-VIADKSG-----VLSGSALEVLQSSTAGTDHGYSSATQVELEPNVTYLSG 1365  
QY 486 FVKKII---OTPYDSEFKDDEKFIQY---RYSLIGSQSHVTLKIBERHYSATQLLNS-- 537  
Db 1366 KIKTDLAKRAYFNIDLRDKDKRIQWIHNEYSALAGKNDW---KROITFTPANAGK 1421  
QY 538 ---TLFQNTDKSELGL-LKQETCTGKNGKTYVVKFTYTKQDDTLQOS-HSITHD 592  
Db 1422 AVVMEVDHKDKGKAWFDEQVLEKEGYSSSNYPVQNSFTSATENMNVNVCASVDSEE 1481  
QY 593 NFTIHRQVRSRYTGRFLFSDTDTXDIIVTQMSYDKLGRLLTET-----L 635  
Db 1482 GFNDVSLKAARTSASQAGSVTKQTVVLGQSDNKPVYLTGMSKASSVKFTDEKDYSL 1541  
QY 636 NSGTPYANTLYDYEL---NNLDQNRPPFVI----- 664  
Db 1542 QANVTYADGSTGIYNAKPPSCTQBNRAAVVVPKTPINKVDISILFKQSATGTWVPDDI 1601  
QY 665 -----TTTVDNGQLRNEFDGAGRHSVQCLKSDGCKEYTHITQYDQGRHHT 714  
Db 1602 RLIEGSLTKSYDSNGYVTKEDELGYATS---TDYDETK---KTSETAKGKXT 1654  
QY 715 STY---SDVLTNGRQOTDPKVLHSMKSYDNWG-QIANT-----HWSYGYSEKI-- 760  
Db 1655 YTYQADQDLTNWLSNGTSILH-----SYDKEGNEVSKTIRAGADQYKFEYDVMGLVK 1709  
QY 761 TVDPITLTATQLOSNMNVOT---GKEVTTYPSQOPIQITLDFEAGHLQSHCHITLRD 816  
Db 1710 TTDPLGNVLASEYDANSNLTKTISPNGNEV-----SLSYD 1744  
QY 817 GMDVRKETDALGCTIYQYQYNNRNVITLPLDGTIVNRKYAPSTDTLITDIRVNGISL 876  
Db 1745 GTDVRKSKSYNGTEKYIFTYQNGN-----E'SVNKEQN-----TT 1781  
QY 877 GQOTFDGLSRLTQSDQGRVWAYTYTYSAGNQCPSVTITPDGQFIHYQVQPELDDAVLOVA 936  
Db 1782 KKRTFDNKNRULTELDTRGSGQWTYPSDSKLKTP-----SWIH-----G 1821  
QY 937 SNEITQOFSYNPVTGALLKVAEGOSLPIIYPSGRKWMENIDMKVSYLWTLRGLENG 996  
Db 1822 DQKGNQFTYN-----KLDQNIENKUSTSYSDYDEN- 1854  
QY 997 YTDLTGCTIKISRDTHGRVTOIKDSSIKITLNYDDLRHIGSQVTDLATGHMLTTTFEFD 1056  
Db 1855 -----GNVQ-----TFITNGGGTFSYSDERNLVSILHIGDKNGGDIETSEYV- 1898  
QY 1057 GLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLVQLQRTQOISYDSRBNLVKQC 1116  
Db 1899 -----DANGNRRTINSS-----ASGKV-----QYBYGKLNQLVK--- 1927  
QY 1117 DGAECTPKYGHISVTONFTYDIYGNITACHTTFADGTED--HAPFKFANPTDPCQLTEV 1174  
Db 1928 -----ETHEDGTVI---EYTYDGFGRKTV-TTIKDGSSKTVNASFNIMN-----QLTKV 1973  
QY 1175 H-----HHPD-DMPDNI-----RLKYDKAGRVINLTDHNG 1203  
Db 1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLTAIVTKKEDKPFATYKYDEKGNRIQKTVN-G 2032

QY 1204 NTEFTYDTLGRLQNGQSVGYDPLNELRVSKQTTDLDCELY-YRETMVNEVRNGEMIR 1262  
Db 2033 KVTNYFYDG-----DSLNVLYETDADNNVTKSYTYGD-----SGQLLS 2070  
QY 1263 LIRGETTIIAQRASKVLLTGTDSQSVILTSQKQNSQAEAYSAYGK-HKSTANDA---S 1318  
Db 2071 YTECKKYFYHYNAHGDIIAISDSGKTV-----AKYQYDAWGNPTKTEASDEVKDN 2122  
QY 1319 ILGYNGERADPVSGVTHLNGVRSYDPTLMBPHT--PDSLSPFGAGGPNYSYCLGDDPIN 1376  
Db 2123 RYRYAGYQYDEETGLYILWARY--YEPRNGVFLSLDPPGSDGSDLDQNGYAYGNNNPVM 2180  
QY 1377 RSDPSGHL-SWQAWTIGMGIAIGLLITATGGMATAAG---GIAAAIASTSTTALAFCA 1432  
Db 2181 NYVDPGH--W-VMLVNVNAGFA-----AVDGYKAYKSGKGMKGAWAAASNFGPGKIFKG 2231  
QY 1433 LSVTSDITSVISGALEDASPKASSITLGVWSGM-----CAAGLAESAIGKGTKLATHLGAF 1488  
Db 2232 ASRAYKFTK-----KAVKITGHTRHGLNOSIGRNG-----GRGVNLRKALNA- 2273  
QY 1489 AEDGENALLKSTSESSRIKMGVTRSLDRE--IVRNEEQVINKHSGRYTDNFMGKGEQAI 1546  
Db 2274 -----VRSPEKVIKQPNGATKYVKKATVVLNKRKGVITAYG-----SS 2312  
QY 1547 LVHGDQKGLYHTEGK 1563  
Db 2313 RAKGSKHVEHTGKGNK 2329

## RESULT 6

ABU18641  
ID ABU18641 standard; protein; 2334 AA.

XX AC ABU18641;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #4168.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX CS Bacillus anthracis.

XX FN WO20027183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 08-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX XX WPI; 2003-029926/02.

XX DR N-PSDB; ACA22511.

XX XX Claim 25; SEQ ID NO 46565; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2234 AA;

Query Match  
Best Local Similarity 4.3%; Score 386; DB 6; Length 2234;  
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;

QY 21 FTQANNFTSAVSGGVDP--TGLYNFIQITLGHIVG-----NGNL-----57  
DB 748 YSKVGHAESEISHEFVPRFSQDSFGIFGDFYASIPVLNGKVNATNGFMSEKIDITLSG 807  
QY 58 -GPTLPILSYSPLNKTDIGFIGNGLSV--YDRKNSLISL-----TGEN- 102  
DB 808 RPDVSVERTYNSOKKVLFGTGWSGLEERWADGNGNLLISTDGANITFRGDNK 867  
QY 103 -----YKVETDRTVKLOOKLDLNRFEKDLKENCYRIIHKSGDIEVL 145  
DB 868 YQAPGTIYLEIKQVSGGVEIKDKQTV-----TFYKSGDAQGR 905  
QY 146 TGFNNNAFDLVP-----KLLNPAGHAIYIDWNFEATQPLNRIYDDLDGHDPL 196  
DB 906 IEYTKDKYGNNTTVEYDGASLSKVKNASGKBLV--OYDGNKKAAARVI---GPDNKT 959  
QY 197 LNLQYQGLIKTILTFPGQ--KEGYTELRLNQLNSIHNFSIGNENPLTWSFGY---250  
DB 960 ITENYDGLLVSSITPEGVYKYD-----NGVLTISIYDPQHTDAKPVKTSYAVENDR 1013  
QY 251 -----TPIGKNGIL-----GOWITSNAPGLKETVYNNQGHHPQSANLPLVPVTL 301  
DB 1014 LVKVTDLPLGRATLLAYNTGSKVELTTPKGRKTVTYND-----AGNFV-----1057  
QY 302 MKQVPGAGQAPAOEYSVTSHNYVGG-----GSGIWNKKLDNLYGLMTEYN- 348  
DB 1058 -KTVEDVGRNLNLTTSYEYNNANLVKTTPKNQETATYDGNVTSVDEMGEKEFYNK 1116  
QY 349 -YGTSEGRYKDK-----GHDQIVRIERTVNNVHLLTSECKQNGYIQTETAYAYAI 400  
DB 1117 DNGIIRKATDNEDRKTIVAYVGAITEVSGQTDQGAN-----TSSVIHHDQYGNPIETSKELS 1171  
QY 401 IGHNFDSQPSOFOLPKTKTETRSADNSYSRSEITETTFDESNGPLKVIKDKTKIISP 460  
DB 1172 AGNLIQNPS-FEM--NGTEKNVKVDYNNSGSIS-----KDATPAGGLGESSLKITTK 1223  
QY 461 STH--WYYPYPAGVDNCPPEPVGFTFRVKII-----QTPYDSEFKDDP 503

DB 1224 ATNNDWGYIAAQEVTLPEPNTVTLTSGMVKTLVNGAFAFNQVSLNENGAGIDGGWHDTR 1283  
QY 504 EKFIQVRYSLIGSQSHVTLKIEERHYSAATQLLNSTLFQYNTDKSELGRL-LKQTECTKE 562  
DB 1284 HNKVQGTSDWNNQ--VTFKTEQ-----TRKVIYLOVENGGSATSGSAWFKIOLEKGE 1337  
QY 563 NGKTSVWHKFTYTKQ--DDTQQ-SHSITTH-----DNFTIHRSOVRGRYTGRLFS 611  
DB 1338 VSSFPVLNSPEENPDGFPQWVRSCQCHERNVDSDSFTGHSSIVMER-----S 1391  
QY 612 DDTKDKIVTQMSYDKLGRLLTTLHSGTYPANTLYDYELANLQDNRPPFVITTTDVG 671  
DB 1392 EYGPNDI-----GYRNRVILNQKAEVTLTAMSKSENVND-----1428  
QY 672 NQLRNEFDGAGRHSOCLSDGDKFYTHIQDYDEQGRHHTSVSYDLTNGRQOTDP- 730  
DB 1429 -----APDKLSKDYAVLAETYYQD-----TVNNYYSFPGSTNDW 1464  
QY 731 -----DKVHLSMSKSYDNWGOIANTHWSYGVSEKIVTDPTLTATKLOLQNS 777  
DB 1465 NRSAAVIPAKKPIQKIEIFLFRKNKG-----KVPFDDIRLLEGNALIKNE 1511  
QY 778 NNVTGKEVTTTPSQOPIQITLDEAGHLOSCHTLTRGDWRVRKETDAIGQCCTIYQD 837  
DB 1512 YD-NDGNVATY-----DEBQK--NTFTVDASGNKKSETDEKGNTKLYDYN 1555  
QY 838 NNNRVIQITLPGTIVNRKIAPFSTDLITDIRVNGISLGQOTPDGLSRLTQSDGGRW 897  
DB 1556 KDNLLTKVTLKNGTSVNYD-----1576  
QY 898 AVTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTGALLKAV 957  
DB 1577 -----HNGNTEKSVWFGKTKTHKEYVDVKNVTYIDALARRIEN--TYDENANKIKTM 1631  
QY 958 AEGQSLTPIYPSPGRKLKMNINDMKNGSYLTLRGLNGYDITLGTIOKISRDTHGRVTQ 1017  
DB 1632 PNGSILESVDYDRAVWGEK-----RNGKDSFT-----FERDQNGQVTK 1670  
QY 1018 IKD--SSIKTILNYDDIARHIGSQVTDLATGHMTTTFVDFGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTKTYDKADR-----VTS--AT-----DSRGGKID 1702  
QY 1076 IQCSWL-----KTQQLANRIVKLNGVLQRTQEQSYSDSRNLNQ--YKCDGAECPDCKY 1126  
DB 1703 -----NAYHDKANSKTEKLEQVITQGGY--TNKVSYD-YNTLDQNIHVTDGSG- 1748  
QY 1127 GHSIVQNTFTYDIGNITACHTTFADGTHEDHATKF--ANP-----TDFCQLTEVHHT 1177  
DB 1749 -----TYRFDYDQGNV-----RTYTAGNGSGSTFNYDQANKIKOLVWGTNSILLSERYE 1799  
QY 1178 HPDMPDNRLKYDKAG-----RVNITDNHNGNTENFTYDTLGR-----LQ 1217  
DB 1800 YDQSGNRKIKEGAGGKVETNFVYDPIQLLNEVLNPGTTKSYTYDGFGRNRSVKVIE 1859  
QY 1218 NQGS-----VYGVDPNRLVSKOTDLDCEL-----1244  
DB 1860 NGKETKTAATFNEGQNLVKEFNESSLTYDVNGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919  
QY 1245 -----YVRETMVNVNVRNGEMIRLLRTGETI-----IAQ 1274  
DB 1920 AFATKYDEDNRIRKQVNGQVTRYFYDGDSDINPLIYETDNGTVLRQVYVSADARLAMK 1979  
QY 1275 RASKVLLTGTOSQSVI--LTSQKONLSQEAISAYGK-----HKSTANDASILYNGERA 1327  
DB 1980 AQGQTLYYHNPGRGDWAMTNDQKEVATYEDAWGNVLTSDTKGIAAD-NPFGYAGMY 2038  
QY 1328 DPVSQVTHLNGYRSYDPTLMRPHTPD-----SLSPFGAG-----GNPKYSYCLGDPINR 1377  
DB 2039 DKEIGMYL-----IARYNPEHGVFLSDPDPGDEDDPVTNGTYADNPNVMM 2088  
QY 1378 SDPSGHLISQAWTIGMGIAGLLTIATGGAIAAAGGIAAIAAIASTTTTALAFGALSVTS 1437  
DB 2089 TDPGKMAW-----LVPVVVIAGAMVAARFCAKAYAI-----RYGAKYKGAOKVKS 2131







749 GNLTHEKTRPVQQLNRFGQVLDRETGLHY--NLYRFYDPDICKFISGD---PIGLAGG 803

1364 INFYSYCLGDPINRSDPSG 1382  
|||::||  
804 INLYQVA-ENPLSYIDPLG 821

RESULT 9  
ABU19676  
ID ABU19676 standard; protein; 1515 AA.  
XX AC AC  
XX ABU19676;  
XX DT 19-JUN-2003 (first entry)  
XX DE  
XX DE Protein encoded by Prokaryotic essential gene #5203.  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
KW Borrelia cepacia.  
OS  
OS  
PN WC200277183-A2.  
PD 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WC-USO09107.  
PF  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
PI WPI; 2003-029926/02.  
DR N-PSDB; ACA23546.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 47600; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

Db 1225 RDG-----TTGQVVTWHIEPGSFLPLAQETDDGLPILTDQIGRPXTVDEQGRPVWKA 1278  
 QY 1303 AYSAYGK---HKSTANDA-----SILYNGERADPVSGVTHLNGYRSYDPTLMRPH 1351  
 Db 1279 AYLWGLKLLPKVRPANDACGATSIDITLRSQGWADETGLNLYNRY--YFDSQYL 1336  
 QY 1352 TPDLSLPPG-AGGINPVSYCLGDPINRSDPSG 1382  
 Db 1337 SAD---PIGLGEGARTOAY-VHDPESQWIDPLG 1364

## RESULT 10

ABU15135  
 ID ABU15135 standard; protein; 1397 AA.

XX AC ABU15135;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #662.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Escherichia coli.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362999P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA19005.

XX PS Claim 25; SEQ ID NO 43059; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1397 AA;

Query Match 3.9%; Score 342; DB 6; Length 1397;

Best Local Similarity 21.4%; Pred. No. 3.1e-14;  
 Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

QY 418 KTETW-RGADNSYRSEITETTFDESGNPLTKVI-----KDKTKKIIS-----PS----- 461  
 Db 218 RTQTFHRAAGFSGEITGVT-DGAGRHERLVLTQAQRAEAEARQAISGCTEFSAPDPT 276  
 QY 462 -----THWEYPPAGEVDCNCPPEP-----YGFT-----PFVK 488  
 Db 277 LFQYTEYGRDNGIRLSAVMLTHDPEYP-----ENLPAAPLVRYGWTPRGELAAVYDRSNT 331  
 QY 489 KLIQTPYDSEPK-----DDPEKFTQYRSLIGSQSHVTLKIEERHYSATQLLNSTL 539  
 Db 332 QVRSFYDDKYRGVMVAHRHTGRPE--ICRYD--SDGRVTEQLNPAGLSYT----- 379  
 QY 540 FOYNTKSELGRLLKQTEC--TKGNGKTYSVVHKFTYTKODDTLQOQSHSTTHDNFTIH 597  
 Db 380 YQYKDRITITDSLNRREVLTQEGG--LKRVRK-----EHADGSS 421  
 QY 598 RSQVESRYTGRLFSDTDKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657  
 Db 422 QSQFPA--VGLRAQTDAGRTTEYSPVVTGLTRIT--TPGRASAFY----- 468  
 QY 658 NRPPFVITTDVNGNQLRNEFDGAGRHYVSCQLKQSD-----GDGKFYTIHTFOQYDEQGRH 712  
 Db 469 NHHSQLSATGPDGLEIRREYDEWGRLIQETAPDGDITRYRYDNPHSDLPATEDATGSR 528  
 QY 713 HFTSTYSDYLTWGRQQTDPKQVHLSKSYDNWQQLANTHWSYGVSEKITVDPI-TLTATK 771  
 Db 529 KTMWSRY---GQLLSFTDCSGYVTRYDHRFGQVTAHREELGSLQVAYDSRQLIAVK 585  
 QY 772 QIQSNNNVQTKGVTTVTPSQQPIQTLFDEAGHLQSCHTLTRDGMWRVREKETDAIGQ- 830  
 Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDAMGKA 623  
 QY 831 CTI-----YQDYNVNEVITLPGDTIVNRKYAPESTDTLITDIRVNGISLQQTFF- 881  
 Db 624 ICTTGGGLTRSMEXDAAGRVIRLTSENGS-----HTTFRYVDLRLIQTGDFGORTQYH 678  
 QY 882 -DGLSRLTQSDQGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEI 940  
 Db 679 HDLTCKLIRSEGLV-----THWYD-EADRLTHRTVNGET 714  
 QY 941 TQOFSYNPVTGAL--LKAVAEGQSILTPY-YPS-GRLLKXENI-----NDM---KMSY 986  
 Db 715 AERWQYDS-RGWLTDISHISEGHRVTYHYGYDGSGRASEHLTVHPQTNELLMQCHETRH 773  
 QY 987 LWTLRGLNGVYDITGTTQKTSRDTHGRVTOIKDSSIKITLNYDDLNHRHIGSQVTDLATG 1046  
 Db 774 AYNAGLAN--RCIPDSLPVAVENLTYG-----SGWLSGMKLG 808  
 QY 1047 HMLTTVTTF--DGLNREITGRKLDCSSGHTLTIQOQSWLTKTQQLANRIVKLVNGLORTQYS 1104  
 Db 809 D--TFLVSYTRDLRHRETLR-----SFGRYELTAYTPAGQLQSQ--HLNLSLS-DRDYT 858  
 QY 1105 YDSRNLRLQYKCDGAECPTDKVGHISIVTONFTYDIYGNITACHTTTFAD-----GTEDHA 1158  
 Db 859 WNDNGELIR-----ISSPRQ-----TRSYSYSTTGELTGVHTTAANLDIRIPTTDP 906







QY 1330 VSGVTHLNGVRSYDPTLMRPHTPDSLPFG-AGGINPYSYCLGDPINRSDPSGH-SWOA 1388  
 Db 1199 ESGLYY-NRHRYYDPLQGRVITQD---PIGLKGNWLYGYQL-NPISDIDPLGLSMWED 1252  
 QY 1389 WTGIGMGIAGLLLTATGGMIAAAGGIAAIAASTSTTALAFGLS VTS DITSIVSGALE 1448  
 Db 1353 -----AKSG---ACTNGLCGTLSA-----MIGPKFDSIDSTAY 1283  
 QY 1449 DASPKASSILGWSMGMAAGLAESAIAKGGTKLATHLGAFAEDGENALLKSTSESSRIKW 1508  
 Db 1284 DALNKINS-----QSICEDKEFA-----1301  
 QY 1509 GVTRSLDREIVRNEGQVIKHSRGYTDNFMWGEGQ-----AILVHGDKDGLFHTGEG 1561  
 Db 1302 -----GLICKDNGRYFSTAPNREGKRGKSYFPNPGNGTEKVSAYHTHG 1346  
 QY 1562 NKHNGK-GPYRTRHTPEQLVDYLDKNNI 1587  
 Db 1347 ADHGEYWDIEFSGDEKIVK-SKNNI 1373

## RESULT 14

AAAB15983  
 ID AAB15983 standard; protein; 1426 AA.

AC AAB15983;

XX 05-OCT-2000 (first entry)

DE E. coli proliferation associated protein sequence SEQ ID NO:340.

XX Escherichia coli; E. coli; proliferation; inhibition; screening;  
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.

XX Escherichia coli.

XX WO200044906-A2.

XX 03-AUG-2000.

XX 27-JAN-2000; 2000WO-US002200.

XX 27-JAN-1999; 99US-0117405P.

XX (ELIT-) ELITRA PHARM INC.

XX Zyskind J, Ohlsen KI, Trawick J, Forsyth RA, Froelich JM;  
 PI Carr GJ, Yamamoto RT, Xu HH;

XX WPI; 2000-514822/46.

XX N-PSDB; AAA65988.

XX Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy.

XX Claim 11; Page 253-256; 316pp; English.

XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
 CC sequences derived from *Escherichia coli* which inhibit *E. coli*  
 CC proliferation. AAA65890 to AAA66055 and AAA65886 to AAA66040 represent  
 CC nucleotide and protein sequences associated with *E. coli* proliferation.  
 CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation- required gene in a microorganism, by contacting  
 CC a microorganism with a proliferation-required gene activity inhibitory  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria

SQL Sequence 1426 AA;

Query Match 3.8%; Score 339; DB 3; Length 1426;

Best Local Similarity 20.2%; Pred. No. 5.2e-14;

Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;

QY 572 KFTYTKDDTTLQOQSHSIITHDNFTIHRSQVRSRYTGRSLFSDTDTKDIVTQMSYDKLGRLL 631  
 Db 317 RYTYTEAGELL-----AVYDRNTQVRAFTYDAQHPGRVMAHRYAGRPMEYRYDDTGRVV 372  
 QY 632 TRTLNMGTFYANTLTYDYELANLQDNRPPFVITTTD-----668  
 Db 373 EQLNPAG-----LSRY-----IVTDSLNRREVLHTEGGAGLRVVKKE 418  
 QY 669 -VNGNOLNEDFGAGRHVSQCLKSDGDGKPYTHITQOYD-----EQGRHHSTYSDYL 721  
 Db 419 LADGSVTRSGYDAAGRLTAQ--TDAAGRRTYGLNVVSGDITDITTPDGRHETKFYND-- 474  
 QY 722 TNGRQOT---DPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQNSN 778  
 Db 475 --GNQLTAVVSPD--GLESREYDEPGLV-----SE 502  
 QY 779 NVQTGKEVT--TYTPSQOPIQITLFEAGHLOSCHTLTRDGDWRVRKETDAIGQCTIYQYD 837  
 Db 503 TSSRGETVRYRYDDAHSELPAITTTDATG---STRQMTWSRYGQLLAFTDCSGYQTRYEYD 559  
 QY 838 NVNRVIQITLPGDTIVNRKYAPFSTDTLITDIRVNGISLQGTDFGLSRLTQSODG-GRV 896  
 Db 560 RFGM-----TAVHRE-----EGISL-YRRYDNRGRLLTSVKDAQGRE 595  
 QY 897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ-----925  
 Db 596 TRYENAAAGDL--TAVITPDGNRSETQYDAMGKAVSTTQGLTRSMYDAAGRVISLTNE 653  
 QY 926 -----PELDDAVLQVANSNEITQOFSYNPVTGALLKAVAEQSLTIPIYPSGRKME 976  
 Db 654 NGSHSVFSDALDRLVQOGGDFGRTQRYHYD-LTGKLTQSEDEGLVILWYDESDRTHR 712  
 QY 977 NINDMKMSYLWTLRGLNGYTDLTGTIQKISRDRTHGRVTQIKDSS-----IKTTLNYDDL 1032  
 Db 713 TVNGEPAEQWQY-----DGHGWLTDISHLSEGHRAVAVHYGDDK 751  
 QY 1033 NRHTG--SQVTDLATGHWL-----1049  
 Db 752 GRLTGECQTVENPETGELLWOHETKHAYNEOGLANRVTPDPSLPVEMVLYGSGYLAKMKL 811  
 QY 1050 -TTTVEF--DGLNREIGRKLCDSSGH--TLIDIOQSWLKTQQLANR-----1089  
 Db 812 GGTPLVEYTRDLHRETVRSFGSMAGSNAAYELTSTVTPAGLQSOHLNSLVYDRDYGS 871  
 QY 1090 ----IVKLVGLQRTQYSYDSNRNLNQYKCDGAE-----PTDKYGHST-----1130  
 Db 872 DNGDLVRISGRQ--TREYGSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPPELHPS 930  
 QY 1131 -----VTQN-----FTYDIYGNITACHTTFADG---TEDHATFKFANPTDPCQLTEV 1174  
 Db 931 TLTWVPDNRIAEDAHYVYRDEYGRLETKTRIPAGVIRTDDET-----975  
 QY 1175 HHTHPDMPDNIRLKYDKAGRVINITD-NHGN---TENFTYDTLGR-----1215  
 Db 976 HHYH-----YDSOHLRVFYTRIQHGEPLVESRYLYDPLGRMAKRWRRERDLTG 1025  
 QY 1216 ---LQNGQSGVGYDPLNRLVSKOTDITDCELYR-----1247  
 Db 1026 WMSLSRKEFTVWYGDG-DRITTVQTTTTRIQTQTVYEGSFTPLIRVETENGEREKAQRS 1084  
 QY 1248 --ETMLVNEVANG-----EMIRLL-RTGETIIA-----QQRASKV 1279  
 Db 1085 LAETLQEGSENGHGVVFPAPBLVRLDLRLEBEIRADRVSSERAWLAQCGLTVEQLARQV 1144  
 QY 1280 LLTGTDSQCS-----VILTSKQNLQSOEA--YSAYGKHKSTANDASILGN---1323  
 Db 1145 EPEYTPARKAHLVHCDHRGLPLALISEDGNTAWSAEYDEWGNQNLNEENPHHV--YQPYRL 1202





us-09-889-874a-23.rag

Tue Jul 6 16:41:07 2004

Db 872 DNGDLVRAISGRQ-TREYGYVSATGRLESVTLAPDLDIRIPYATDPAGNRLDPDELHPDS 930  
QY 1131 -----VTQN-----FTYDIYCNITACHTTFADG---TEDHATFKFANPTDPCOLTEV 1174  
Db 931 TLTVWPDNRIAEDAHYVYRDEYGRUTEKTRIPAGVIRUDDERT----- 975  
QY 1175 HHTHPMDNIRLKYKAGRINITD-NHGN-----TENFYDTLGR----- 1215  
Db 976 HHYH-----YDSQHLVYTRIQHGEPLVESGRYLYDPLGRMAKRVWRERDLTG 1025  
QY 1216 ----LONGGSGVYGDPLNLVSQKTDLTDCELYR----- 1247  
Db 1026 WMSLSRPEVTWYGDG-DRLTTVOIDTTTRIQVYEPGFTPLIRVETENGEREKARRS 1084  
QY 1248 --ETMLVNEVRNG-----EMIRLL-RTGETIIA-----OQASKV 1279  
Db 1085 LAETLQEGSGENGHVFPFAELVRLDLRLEEIRADRVSSERAWLAQOGLTVEQLARQV 1144  
QY 1280 LLTCTDSQOS-----VILTSDKQNLQSEA-YSAVCKHKSTANDASILGYN----- 1323  
Db 1145 EPEYTPARKAHLVHCDHRGLPLALISEDGNTAWSAEYDEWGNQNLNEENPHHV--YQPYRL 1202  
QY 1324 -GERADPVSGVTHLGNVGRSYDPTLMEFHTPDSLSPEG-AGGINPYSYCLGDPINKSDPS 1381  
Db 1203 PGQHQDEESGLY--NRHRYVDPLQGRYITQD---PMGLKGGWNLQYPL-NPQQIDPM 1256  
QY 1382 GHLSQWQWTGIGMGIAGLLLTATGGMALAAAGGIAAAIASTTTALAFGALSVTSDITS 1441  
Db 1257 GLL--QWDDARSG-----ACTGGV-----CGVLSRIIGPSKFDSTADALD----- 1296  
QY 1442 IVSGALEDASPKASSILGWVSMGMAAGLAESAIGKGTKLATHLGAPAEDEGENALLKSTS 1501  
Db 1297 ----ALKETQNRS-----LCNDMEYSIGVCKDTNG-----KYFASKAETDNLK--K 1336  
QY 1502 ESSRIKWGVTRSLDREIVRNEEGQVIXDHSRGYTDNFMGKGEQAILVHGDK--DGFLYHT 1559  
Db 1337 ESYFLKRCPCPTGTDORVAAYHTHG---ADSHGDVYVDEFFSSDDKNLVRSKDNLEAFYLAT 1393  
QY 1560 EGNKH---NGKGPY 1570  
Db 1394 PDGRFEALNNKGEY 1407

Search completed: July 3, 2004, 05:46:37  
Job time : 89 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:43:29 ; Search time 28 Seconds  
(without alignments)  
3084.652 Million cell updates/sec

Title: US-09-889-874A-23  
Perfect score: 8879  
Sequence: 1 VVIXFLKFRITWSDNEF.....PRKILGTEKTVKTRFP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/BCTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1638	18.4	1584	3	US-09-251-645-6
2	331.5	3.7	1377	4	US-09-711-164-467
3	317	3.6	998	4	US-09-252-991A-28424
4	313	3.5	1627	4	US-09-328-352-6604
5	306	3.4	1586	4	US-09-543-681A-5329
6	294	3.3	1439	4	US-09-543-681A-7560
7	292	3.3	1626	4	US-09-252-991A-23805
8	285.5	3.2	1596	4	US-09-328-352-5542
9	277.5	3.1	974	4	US-09-252-991A-23640
10	271.5	3.1	1665	4	US-09-543-681A-4476
11	238.5	2.7	1043	4	US-08-851-567B-61
12	235.5	2.7	804	4	US-09-328-352-5545
13	223	2.5	1128	4	US-09-252-991A-31032
14	222.5	2.5	3290	4	US-09-328-352-5486
15	220.5	2.5	1183	2	US-08-447-031A-2
16	218.5	2.5	2123	3	US-08-968-685A-10
17	214	2.4	2504	4	US-09-328-352-5821
18	212.5	2.4	954	3	US-09-251-645-12
19	211.5	2.4	10182	4	US-09-134-001C-3159
20	207	2.3	2314	4	US-09-268-347-49
21	202.5	2.3	2777	4	US-09-543-681A-6124
22	201	2.3	1739	4	US-09-540-236-3739
23	201	2.3	2057	4	US-09-489-203-2
24	198.5	2.2	1084	4	US-08-268-347-30
25	190	2.1	2385	4	US-09-543-681A-6304
26	188.5	2.1	2736	4	US-09-252-991A-30227
27	185.5	2.1	1166	4	US-09-200-650E-7

28	184.5	2.1	2142	4	US-09-540-236-3459	Sequence 3459, Ap
29	184	2.1	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
30	183	2.1	1222	4	US-09-206-942-37	Sequence 37, Appl
31	183	2.1	1228	4	US-09-206-942-34	Sequence 34, Appl
32	182.5	2.1	1565	4	US-08-851-567B-59	Sequence 59, Appl
33	182	2.0	1861	2	US-08-790-912-4	Sequence 4, Appl
34	181.5	2.0	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
35	181	2.0	1228	4	US-09-463-402-2	Sequence 2, Appl
36	181	2.0	1228	4	US-09-889-572-2	Sequence 2, Appl
37	180.5	2.0	930	4	US-09-134-001C-5314	Sequence 5314, Ap
38	180.5	2.0	930	4	US-09-386-962C-10	Sequence 10, Appl
39	180	2.0	1683	3	US-08-755-587-183	Sequence 183, Appl
40	179.5	2.0	1833	4	US-08-621-944A-4	Sequence 4, Appl
41	179.5	2.0	1833	4	US-08-945-567D-4	Sequence 4, Appl
42	179.5	2.0	1992	4	US-08-621-944A-3	Sequence 3, Appl
43	179.5	2.0	1992	4	US-08-945-567D-3	Sequence 3, Appl
44	178.5	2.0	1338	1	US-08-471-033-50	Sequence 50, Appl
45	178.5	2.0	1338	2	US-08-471-044-50	Sequence 50, Appl

## ALIGNMENTS

## RESULT 1

US-09-251-645-6  
; Sequence 6, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1584  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-6

Query Match 18.4%; Score 1638; DB 3; Length 1584;  
Best Local Similarity 31.0%; Pred. No. 9.5e-118;  
Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;

QY	19	EFTQANNFTSAVSGVDPRLGLNIQITLGHVGN--GNLGPTLPLTSLVSPLNKTDIG	76
DB	5	DIYSNAPFGSYINTGVDPRTGQYSANIITLRPNVNGSEQT--LSLSFPLTLNNG	62
QY	77	FGIGNFGLSVYDRKNSLSLSTGENYK--VIEDTKVTKQKKLDNLRFEKOLKENCY	133
DB	63	FGIGNRFSLTLDIKTLTFSRANGEQFKCKPLPNNDLSFKDKLKDRLRYK-LDSWTF	121
QY	134	RIIHKSQIEVLGTGNNAFADLVKPKLNPAGHAIVYDWNFEATQPRLNRIYDLDGDH	193
DB	122	XYVKNKGIEILKRIGSS--DIAKTVALEFPDGEAFDLIYNSR-----	162
QY	194	IPLINLEYCGLIKTLTLT--FGQKEGYRTELRFL-NQLNSIHNFSLGNENPLTWSFGYT	251
DB	163	FALSEIKRVTGKTYKLKLYSGNN---CTSVEYDDNNISAKIAFDVRNLIITVTVPYD	219
QY	252	PIGK-----NGILQWITSMTAPGLKETVYNSNNQGHHPQSANLPVLVY	298
DB	220	ASGPIDSARFKMTYOTLKGVF-PVISTFRPTGTVELSVYKEN--GH---KVITETIPIY	273
QY	299	VTLMKQVPGACQAIQAEYSYTS-HNYVGGSGNCI-WNNKLDNLYGLMTEYNGSTESRR	356



QY 644 TLTYDYELNLODDNRPFFVITTTDVGNGQIARNBPDGAGRHVSQCLKDS-----GDKF 698  
Db 463 ASAFYNHNO-----LTSATGPDGLERREYDELRLIQTETAPDGDITRYRYNPH 514  
QY 699 YTIHQYDEGRHHTSYSTYLTNGROQTDPDKVHLSMSKSYDNWGQIANHWSYGVSE 758  
Db 515 SDLPATEDATGSRKMTWSRY---GQLLSFTDCSGYVTRVDHDFGQMTAVHREGLSQ 571  
QY 759 KITVDPI-TLATKQLOSNVNVQTKVITVTPSQPIQITLDFEAGHLQ-----SC 810  
Db 572 YRAVDSRGLIAVKDQGHETRYE-----YNIAGDLTAVIAPDQSR 612  
QY 811 HLTTRDGDWRVRKETDAIGCTIYQYDNNRVIOTLTPDGTIVNRKYAPSTDLITDIR 870  
Db 613 NGTOVDANGKAVRTTQ--GLTRSEYDAAGRVILTSNGS-----HTTRYDVLRLIQ 666  
QY 871 VNGISLGQTF--DGLSRLTQSQGGRWAYTYSAGNDQCPSTVITPDGFIHYQYQPEL 928  
Db 667 ETGPDGRTQRYHDLTGKLIASEDEGLV-----THWYD-EA 702  
QY 929 DDVLOVASNEITQOFSNPNVTGAL--LKAVAEQSLTPIY---PSGRL--KMNENDM 981  
Db 703 DRLTHRTVKGTAEARWQDE--RGWLTDSHISEGHRVA--VHYRYDEKRLTGERQTVHP 760  
QY 982 KMSYLM-----TLRGLNGYDITLGTIYQIKISRDTHGRVTQIKDSSIKTILNYDDL 1033  
Db 761 QTEALLWCHETRHAYNAQGLAN--RCIPDSLPAVEWLTG-----SGYLAGMKLD-- 809  
QY 1034 RHIGSQVDTLATHMLTTFEF--DGLNREIGRLKLDSSGHTLDIQOSWLKQTOANRIV 1091  
Db 810 -----TFLVEYTRDLRHRETLR-----SFGVELTAYTPAGLOSO-- 846  
QY 1092 KLNGVLTQEOYSYDSNRNLNOYKCDGAECPTRKYGHSIVTONFTYDIYGNITACHTTPA 1151  
Db 847 HLNSLLS--DRDYTWNDGELIR-----ISSPRO-----TRSYSTTGRITGVHTAA 893  
QY 1152 DGTEDHATFKPANDTDC--OLTEVHTHPD-----MPDN-----IRLKYDAGRVIN 1197  
Db 894 -----NLDIRIPYATDPAGNRLPD--PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTE 947  
QY 1198 ITD-----NHGNTENTYDTLGLRONGQ-----GSVGYDPLNRLVSKTDT 1239  
Db 948 KTDLIPGVTITDDERTHRHYDSQHLVHYTQVEEPLVESRYLYDPIGRVAKRVWR 1007  
QY 1240 LDCBL-----YY-----RETMVNE-----VRNGEMIRLLR-----TGETI 1270  
Db 1008 RERDLTGWMSLSRKPQVTWYGDGDRLLTQNDRTRIQTIVQPGSFTPLIRVETATGELA 1067  
QY 1271 IAOORA-----SKVLLTGTDSQOSVIL-----TS 1294  
Db 1068 KTORSLADALQOSGEDGSGVPPVVLVQMLRLESEILADRVSPESRRWLASCGLTVE 1127  
QY 1295 DKQN-----LSEOA-----YSAYGKHKSTAND--AS 1318  
Db 1128 QMOMQDPVYTPARKIHLXCHDRGLPLALISKEGTENCAEYDEWGNLLNEENPHLOQ 1187  
QY 1319 ILGNGERADPVSGVTHLNGYSYDPTLMRFHTPDLSLSPFG--AGINPYSYCLGPPINR 1377  
Db 1188 LIRLPGQYDEBSGLAY--NRHRYDPLQGRYITQD---PIGLKGMNFYQYPL-NPVTN 1241  
QY 1378 SDPSG 1382  
Db 1242 TDLPG 1246

RESULT 3  
US-09-252-991A-28424  
; Sequence 28424, Appication US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28424  
; LENGTH: 998  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28424  
  
Query Match 3.6%; Score 317; DB 4; Length 998;  
Best Local Similarity 22.0%; Pred. No. 2.2e-15;  
Matches 222; Conservative 129; Mismatches 362; Indels 296; Gaps 47;  
  
QY 504 ERFQIYRSLIGSQSHVTLKIERHYSATQLNLSILFYNTDKSELGRLLKOTECTKGN 563  
Db 112 ERAWSYSNALG-----LIBRADGPRTDVQDVTLYAYD-----SRGNL 149  
QY 564 GKTYSVVKHFTYTKQDDTLQQSHSITTHDNFTHSQVRSRYTG--RLFSDDTDDIVQ 621  
Db 150 TQVTNALGOVTRLGDYDERGKPSITDANGVTSSLA-----YTGVDGMLASVSTAGSTR 204  
QY 622 MSYDKLGRLLTSLNSGTPYANTLYDYELNNLQDNRPPFVITTTDVNGNQLRNEFDGA 681  
Db 205 PDYDVGQYTRVTRGDG---SWLSYEV-----DARR--LVAATGNLGERLEYDVTK 252  
QY 682 GRHVSQCLKSDGDKGKFTIHTQO--YDBQGR-----HHTSYS--DYLNGRQOQDP 730  
Db 253 GNRTAQRIKASGS-----LVRCQWAYDELGLRLRAVGAGGQTRFAVDLNDNPVGETNP 308  
QY 731 DKVHLSMSKSYDNWGQIANHWSYGVSEKITVDP--TLTATKQLOSNVNVQTKVITV 790  
Db 309 RC--FAHSQAFDALDLVQSDPLGKTRLAYD-----AODNLTEVKDPRGVTRY 357  
QY 791 PSQPIQITLDFEAGHL-----QSCHTLFDGWRVKETDAICQCTIYQYDNNRVI 843  
Db 358 E-----YDGLNLRILVSPDSGTTTFHDAAGNVIRRTDAEGAVTEYRYDALNRLV 408  
QY 844 QITLPDGTIVNRKYAPSTDTLITDIRVNGISLGQTFDGLSRLTQSQDQGRVWAYTSA 903  
Db 409 ERRSP-----SDPSLDVQYRYDLTADGNGQIGRLGAIENGADS-----446  
QY 904 GNDQCPSTVITPDGQFIHYQYQPELDDAVLQVANSNEITQOFSYNPVTGALLKAVAEQSL 963  
Db 447 -----LVRYD--ERGNLVEQVRSIRLDQQTLLDRVTRYDAA-----NQL 485  
QY 964 TPIYVPSGRLKMNENDMKMSYLTWLRGLENGYDTLGTIYQIKISRDTHGRVTOIK-- 1019  
Db 486 LEIGYPS-----GLAIGY-----PRNAGGQVASTLAVG 514  
QY 1020 DSSIHTT-----LNYDDLNRHISQVTDLATGMLTTFVFDG-----LNREIGRLKCLS 1069  
Db 515 DKAPSTLVQIAYLFPFGPLQR-----LTWNGITLSREYDQDYQLLRQKVG-----560  
QY 1070 SGHTLDIQOSWLKQTOANRIVKLVNGVLR-----TEQYSYDSRNLNOKY--CDGAE 1121  
Db 561 -----PWQSDYQ-----HDANGNIQQRHSLWGLTDYQYDPLDLTEERGQVGG-- 604  
QY 1122 PTDKYGHSIVTONFTYDIYGNIT--ACHTTFADGTEDHATFKFANPTPCQLTEVHTHPD 1180  
Db 605 -----RSYADVAGNRTQSDNPASGGTASSQDYQA-----636  
QY 1181 MPDNRL-----KYDKAGRVNITDNGHTNENFTYDILGRON-----GQSVTVGYD 1227  
Db 637 -PDNRLTGAIGAQAVTSDAAG---NLQDRA--ARKLAYDAQGRQLQSVLSLQQAERYN 691  
QY 1228 PL-NRLVSKQTDITLDCELYVRETMVNEVRNGEMIRLIRTG-----ETIIAQORA 1276  
Db 692 ALGRIVVKLTESITTYLYGPDGQLLGEAHDGSRKLRQAYYLWLSLPLATTDADYDA 751

QY 1277 S-----KVLITGTSQQSVILTSK--QNLSEAYSAYGKHKSTANDAS--ILYNGER 1326  
 Db 752 QGKVGNTLLYLHGDHLDTPRLATDASGQIAWQWSDAFGRGEALSQSGSTQVNLRFPGQY 811  
 QY 1327 ADPVSGVTHLNGCYRSYDYLRFHTPDSLSFPG-AGGINPYSYCLGPPINRSDPSG--- 1382  
 Db 812 YDAESGLHY--NYFRYDDETGRYVESD---PIGUSGGVNTYGVYQGAFLNRIDPLGLAA 866  
 QY 1383 ---HLSQWAWTIGMGI---AGLLTIATGMAIAAGGIAAIAASTST 1425  
 Db 867 IEIDIPKSAVDMPGNRLPACGLL---GGVLLVASISGATPQADSPT 911

## RESULT 4

US-09-328-352-6604  
 ; Sequence 6604, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604  
 ; LENGTH: 1627  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6604

Query Match 3.5%; Score 313; DB 4; Length 1627;  
 Best Local Similarity 20.4%; Pred. No. 1.1e-14;  
 Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;

QY 75 IGGIGFNGSLVYDRKNSLSLSTGENY--KVIETDKTV-----KIQ 115  
 Db 369 IFAMTEFTFHDVQALQGIQDSIRYVSNLYOMDDAIFGARWVPTTKISKRFY 428  
 QY 116 QKLDNLRPEKDKENCYRIIHKSGDIEVLTFGNNAFDLPVKKLNPAHAIY----ID 172  
 Db 429 SKKCHKYKLVNGLEVIC-----LDGRAIDLPLKK-----GQSIYDPIEQ 468  
 QY 173 WNFETQPLNRI-----YDDLGDHDIPLLNLEYQGLKLTILTFPGQKGYRTEAR 224  
 Db 469 YTTVLSDQLHLIAYGEDEKRYEKGEDYRLSYIE-----RKGFKVALR 514  
 QY 225 F-----LNRQLNSIHNFLSGHENPLTWSFGVTPIG-----KNGILGOWITSMTPGG 271  
 Db 515 YDHVSTDRKTLISDLIFKQDDNLLAHLALQLTPQGLVSDIMTWIKNGQLDRVLAS----- 568  
 QY 272 LKETVYNNNGHFFPQSANLPVLYTLMKQVPGAGQPAIOAEYSVTSH-----NYV 325  
 Db 569 -----YDYOQ-----DLVQATNEPFAASYIYQYTHLITRYDLT 604  
 QY 326 GGGNGIWNKLDNLYGLMTEYNGSTBSRYKQEGHDQIVRIERTYNNYHLLTSECKQ 385  
 Db 605 HRGMNKKWDGILPTSKAIEWADNASRASKLEWDK-----NIRKT-----TVLDVEGNS 653  
 QY 386 QNGYIQTTEYAYIIGHNFDSQPSQFOLPKTKETWRSADNSVRSBITETTPDESQNP 445  
 Db 654 TEHYVIDIGYTRIVYDNPFE---ECFPRDDAKNITHIAKDGSK-----TSYTVDERGNVL 707  
 QY 446 TKVIKDKTKQKISPSSTHWEYYP-----AGEVDNCPPEPY-----GTFRFVKIITQPYDSE 498  
 Db 708 TTTQDDGAT-----SYFEYDEKNQLTGMVDAEQGRWFKQYDQSGNLIKEI----- 752  
 QY 499 FKDDPEKFTQYRYSLSIGSQSVHTLKIERRYHYSATOLLNSTLFOYNTDKS-----ELGRLLK 554  
 Db 753 ---DPLK-----HETAYV-----YVAGLVTSITDAKGGSKSLKYDDQGNLIS 792  
 QY 555 QTEC-----TK-----GENGKTYSVVH-----KFTYTKQDDTLQO----- 594

Db 793 YTDCSGKTKQYDERGVSIGIENALNQKVBYFYFTELTLENREPIIKGLPLNAFGQLEKI 852  
 QY 585 SHSITTHNFTHRSOVSRYTGLFSDTDKDVTCMSYDKLGRLLTRT--LNSGTPYA 642  
 Db 853 KHAQGTBEHF-IHRAE-----GRLLAHVDPKQNIYRYEDEAGLILSRDALN----- 899  
 QY 643 NTLTYDYE-----LNNLQDNRPFPVITTTDVNGNQLNEEDFGAGRHVSQCLKSDSDGKRF 698  
 Db 900 HKLKYKWRDLRGLRLTELNE-----NGASYQFFYDVASRLV-----KEIDFDGKE 943  
 QY 699 YTIHTQQYDEO-GRHTS--TYSDYLTNGROOTDP-DKVHLSMSKSYDNMGQJANTHWSY 754  
 Db 944 TVYH---YDEKSGQLATSIEVASYGQDLKDRAPKDRIOQFI---FDSMGRLEQRTAGY 997  
 QY 755 G-----VSEKLTVDPI--TLTATKOLQSNNVQTKKEVITYTTFPSQOPIQITLFDRA 804  
 Db 998 GHYGLEEEKQTEFAVDMGRIITQAKNAOSNLQ-----WFDAA 1037  
 QY 805 GHLQSCHTLTRDGDWRVRKETDAIGQCTI--YOYDYNRVIOITLDPDTIVNEKYAPFST 862  
 Db 1038 GNLVQEH-----QQDYKINKTAVWKHQYDEINDEIKTRPDGQVID----- 1078  
 QY 863 DTLITDIRVNGISLQOQTFDGLSKLTQSDGGRVWYATYSAGNDQCPSTVITPDGOFIHY 922  
 Db 1079 -----W-LTYGSGHVQ-----SLVNGQDFVSF 1100  
 QY 923 QYQPELDDAVLQVASNITQOFSYNPVTGALLKAVAGQSILPIYPSGLKVENINDMK 982  
 Db 1101 -----ERDDLHREIA-----RHYANGVSQEQOYDLAQLKSKQM----- 1134  
 QY 983 KMSYLWLTLGLENGYTD-----LTGTIQLKISR-----DTHGRVTOIKDSSI-KTLLNY 1029  
 Db 1135 -----LSEHENGYQNYKRRHNNALQTSQVQLVQYDKTGELTAIRDRGNIAKY 1187  
 QY 1030 DDLNRHIGSQVTDLATGMLTTFVEFDGLNREIGRKLCDSSGHTLDIQOSWLKTOOLA-- 1087  
 Db 1188 DPGVRLLE-----ASSKLKETFSFDPASNIL-----DSYHSQVQSHSQKLDTSYG 1235  
 QY 1088 -NRIVKLVGLQR--TSQYSYDSNRLNLYKCDGACGPTDKYGHSIVTONFTYDIYGNIT 1144  
 Db 1236 YNRLV--NNVVKEYLDQOYQYDAYGQIROKTSQGL-----NLEMDVYGRMV 1281  
 QY 1145 ACHTTFADGTEHATFANPTDPCQLTEVHTHTPMDPNIRLKYDKAGRVINITDNHGN 1204  
 Db 1282 KSRNS-----QYTAEY-----RYDALGRRIQKWSKHH 1309  
 QY 1205 TENFTYTLGLQNGQSVGYDPLNRLVSKQTDITDCELYYEETMLVNEVRNGEM--IR 1262  
 Db 1310 T-----GOEQN---IITYGWDG-DTLAYESTEELTKHYIYEKDSFVPMLOAVLSPIE 1357  
 QY 1263 LLRTGETI-----IAOQRASKVLLTGTDSQSVILTSKONLSQE-----AYS 1305  
 Db 1358 LHQTPWSDRPNYHHRDPLWKTEKEGEPDDWFWFCHDLGTQEMWTDHTGAILWKAERY 1417  
 QY 1306 AYGHKXSTANDASI-----LYNGBRADPVSGVTHLNGVYSYDPTLMRHHPTDS 1355  
 Db 1418 AWGECKAEEAKSNFENSEIISNNIRFQGYQFDEETGLHY--NRYRYYSFYGRFVSKD- 1474  
 QY 1356 LSPFG-AGGINPYSYCLDPPINRSDPSGHLWSQAWTGIGMGIAQ 1398  
 Db 1475 --PGLLGGNNVYVA--KNPITWIDSKGLSTLNRNLG-GVKG 1514

## RESULT 5

US-09-543-681A-5329  
 ; Sequence 5329, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5329  
; LENGTH: 1586  
; TYPE: PR  
; ORGANISM: Proteus mirabilis  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (1576)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-543-681A-5329

Query Match 3.4%; Score 306; DB 4; Length 1586;  
Best Local Similarity 22.1%; Pred. No. 3.6e-14;  
Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

QY 499 FKDDPEK-----IQYYS-----LIGSQSHVTLKIEERHY-SATQLLNS- 537  
DB 486 FEANPHKINTLRLMKSADRHNNALHYRYVANDGELVQIHDDAVLTDIRLHYDEITQRLQSV 545  
QY 538 -----TLFQYNTDKSELGRLLKQTECKTGKNGKTYSVVHKFTYTKQDDTL---QQS 585  
DB 546 TRHQQOEKTLVITYYDAQ--RLVQVTNADK-----RVTRBFGWDDSEGLMAMQYA 596  
QY 586 HSITH-----DNFTIHRQVRSRYTGLRFSDTDTKDIVTQMSYDKLGRLLTRTLNSGT 639  
DB 597 TGVSSHYRQWRFDAFTIEDNEPEWVVEHMLKDGKRCLEHTELYDLAQRLLTVETGGE 656  
QY 640 PYANTLTYDELNLQDNRPFFVITTTDVNGNQLRNEFDGAGRHVSQCLKDS---DDGD 696  
DB 657 -----TTFRRNEQQIIEYTNALNETWPEWD-----TSRLKKAIPDGE 699  
QY 697 KFYTIHQOYDEQGRHSTYSYDLTNGRQOTDPDKVHLSKMSKYDWMGQ-----IANT- 750  
DB 700 WGYT-----YBERG-----NLQWTDSEQ-----QSTCYD-WDKDFAPPTAQL 737  
QY 751 -----HWSYGVSEKI--TVDP-----ITLTAIKOLQSNNNVQTKGVITYTTPSQOQIT 799  
DB 738 PNGAAMHWEYNEHGDRIIRVIDPLGHITRLAWDDQGLGQVDAKGNETHRYNARGOLIE 797  
QY 800 LFDGAGHLQSCHLTRQGDWRVREKETAIGCTIYQYDYNVRVLIQITLPGTIVNRKYP 859  
DB 798 QRDSCGPT-----TUTYDDWGQLRSLTNAQNETTTTTFSEAGLLTECLPDT--ENRYDY 852  
QY 860 FSDTL--ITDIRVNGISL--GQ-----QTFDGLSRLTQSQDQGRV 896  
DB 853 DATQGLVITDAGERHILLRNRREGVIAERDPAGHMLHFHYDTFGHMQALENEQ--GEQ 910  
QY 897 WAYTYSAGNDCPSVTITPD--GQFIHYQY-----PELDDAVLQVASNE 939  
DB 911 YRFEYDALHR-----LTDEHDLIGQOKHYQYDVMGNVTOIKTTPGPSIDTPI--PLSPQ 962  
QY 940 ITQOFSYNPVTGALLKAVAEQSLTPIYPSGRLLKMNINMKMSYLM-----TLRGLE 994  
DB 963 VT-TFGYDKV-----GRLLFRENADY-RTEVLYOPLSVTLRRVP 999  
QY 995 NGY---TDLTGTQKISR-----THGRVTQ--IKDSSIKTTL--NYDDLNR----- 1034  
DB 1000 MAVWHEAERTGTTARVEYQDALFTYDKVQGVLRASARGDYQHYYDVLGNITRTELPHQ 1059  
QY 1035 -----HIGS---QVTDLATGMLTITVEF--DGLAREIGRKLCDSSGHTLDIQOSWLKT 1083  
DB 1060 RAEFLYGGSHLOOTWRNEQUTVLAIEQORDLHRETLR-----TSG-ALDNETGYDCR 1114  
QY 1084 QQLANRIVK-----LNGVLQKTEQYSYDSNRNL-----NOVK 1115  
DB 1115 GRITHQVARKMASQFVTPVIDR--RYRWDKNAQLIERSVSYQGTGEVFTAGHWYHSYQ 1172  
QY 1116 CDGACECTDYGHISIVTONTYIYGNITACHITTFADGTEDHATFPAFNPDPCLQTEVH 1175

DB 1173 YDLGQLTAHLG-SVQTEHFLYDAANLL-----TRPH-----TEAP 1208  
QY 1176 HTHPDMPDNIRLYKDKAGRVINITDN-----HGNTEN-----ET 1209  
DB 1209 HNVQGSQDKFDYRVDGFRMVSVYKSGSSGQRYHYDSHRIIAVDDIDQDQPLGYQRAEYR 1268  
QY 1210 YDTLGR-----QNGQSVYGYDPLN-RLVSQKTDITLDCELYRETW 1250  
DB 1269 YDILGRRIEKRLKASAIANTVTYHQHEPDDEVYTFGVMGRVLVSEHSSAAPHTTVYH-- 1325  
QY 1251 LVNEVRNGEMIRLLRTGETIIAQORASKVLLTGTDSQSVILTSQKQLNSQEAYSVAG-- 1308  
DB 1326 AYNDQSVTPLEARIECT-DNPLNFORAIYTHSHLSGLPEALTNSEGEIVWQGYSAWHL 1384  
QY 1309 -----KHKSTANDASILGYNGERADPVGSVTHLNGYRSYDPTLMRPHTPDLSLSPFG-AGG 1363  
DB 1385 QRCRTPTSTFRNQNLRFQGYFDKETGLHY--NTRYIYAPDLGRFTQOD---PIGLAGG 1439  
QY 1364 INPVSYCLGDPINRSDPSG 1382  
DB 1440 INLYAYA-FNPLTWDPWG 1457

## RESULT 6

US-09-543-681A-7560  
; Sequence 7560, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:

; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCES: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7560  
; LENGTH: 1439  
; TYPE: PR

; ORGANISM: Proteus mirabilis

US-09-543-681A-7560

## Query Match

Best Local Similarity 20.3%; Pred. No. 2.6e-13;

Matches 251; Conservative 141; Mismatches 386; Indels 460; Gaps 62;

QY 347 YNYGTESTERYK-----DKEGHDQIVRIERTYNNYHL-----LTSECKQONGYIQ 391  
DB 376 YHYPFDDKGYSRSLSHVDNVEQORFH---YNEHHQLIHITGCDLNECEYQS--FQ 429  
QY 392 TTE-----TAYYAIIGHNFDSPSQF-----QLPKTKTET-----WR 423  
DB 430 LAEKTVSRLTAYQVNPQHRIIRLCAFYNESQAQLIRVEQQTNHPYRQFGWTAGVMAWH 489  
QY 424 SADNYSRREITETTTDESGNPLTKVKKKQKIISPSHWEYYPAGEVNCPEPYGF 483  
DB 490 SKYGLRSEYRWALSEDN---LWRVIENTKTS-----GESYRLEYDDINL 531  
QY 484 TRFVKKIITQPYDSFEDKDDPEKFIQYRYSGLSQSHVTLKIEERHYSATQL----- 535  
DB 532 TR-----TAY---WHQGSTFWQLNH-----DHQIHHYIDRTGIKTLIDBFGPLC 575  
QY 536 ---NTLFQYNTDKSELGRLLKQTECKTGKNGKTYSVVHKFTYTKQDDTLQOSSHITPHD 592  
DB 576 GCRNALGHTH-SEWDALGRLLSITD---GNGNQT-----RWQYQNERERL-----IT--- 619  
QY 593 NFTIHRQVRSRYTGLRFSDTDTKDIVTQMSYDKLGRLLITRLNSGTPYANTLYDYELN 652  
DB 620 -----VFPDNT-----SRLAYSLGRLIKEI-----SPHQITRYRDFK 656  
QY 653 NLQDNRPFFVITTTDVNGNQLRNEFDGAGRHVSQCLKSDGDKFYTHITQQYDEQGRH 712

657 TTL---RP-----TIRIDAKQSRSEFLWNR--GOLLRHTDCSGK---QHIMCYDDEGRV 703  
QY 713 HTSYSDYLTNGRQOQTDPKVHLSMSKSYDNGQIANTHWSYGVSEKITVDPITLTATKO 772  
Db 704 VSO-----FNALQE-----ATEYQYDEVGHARI-----ILPD-----NSTVQ 736  
QY 773 LOSNSNNVQKGVTTTYSQPIQITLFEAGHLQSCHTLPRDGDWRVRKXETDAIGQCT 832  
Db 737 LAMNAAGLTHHQNDNTPCQOY-----NAFGRVTTTIDKLARI 777  
QY 833 IYQYDNNRVIOITLPG--TINRKYAPSTDTLITDIRVNGISLGQOTFDGSLRTQS 890  
Db 778 HYHNAAGALISNANGRYLNRD-----ASRLVEELRDE--TLQYTTINAGRLVEE 832  
QY 891 QD--GGRVWA-----YYSAGNDQCPSTVITPDQOFIHYQYQPELDDAVLQVASNEITQ 942  
Db 833 AHLGDRVTSAPRTILLDYDAAGNLVKRETLTD-----RYQYQWDSMNRL--VASKQBNQ 886  
QY 943 -----QFSYNPVTGALLKAVAGOSLPIYPS 970  
Db 887 RLEMLGQANQVHTVDALRIIREQTDGDIIEFVYDNLNLSRLTLPGQSLANLYGS 946  
QY 971 GRKLMENINDMKMSYLMTLRLGLENGYDITLQTIQISRDTHGRVTOIKDSIKITLNYD 1030  
Db 947 GHATAIN-----HLVDSR-----SOLITEPERD 969  
QY 1031 DIARHIGSQVTDLATGHMLTTTVERDGLNREIGRKLCDSSGHTLD---IQOSWLKTQOLA 1087  
Db 970 DLHREIS-----RTQELTQYQYDKUGRTISTSSRDQKQPLNGITLWRKWFYDQ-- 1021  
QY 1088 NRIVKLVGLQRTQYQYSDSRNLN-----QYKCDGAECP 1123  
Db 1022 GNLGAMEDTVRGVVEYLYDSEQLKKAASSENLDAMLFYDRADNILLRPERQSEMDEHSPT 1081  
QY 1124 -----DKYGHISIVTQNTYDIYNIPTACHTTFADGTEDHATKFAFNPDPCCOLTEVHH 1176  
Db 1082 LELSPQDKL--ROFQWHYQYDAYGNVIA-----RKYRN-----OSSQYI-- 1120  
QY 1177 THPDMPDNLRYDKAGRVINITDNGNTNFNTYDTLGR-----RKYRN-----OSSQYI-- 1215  
Db 1121 -----AYDGNRLV--IAHQGIKAQVHYDALGRRIHKTVENRESQVXRQETHF 1168  
QY 1216 -----LQNCQGSVY-----GYDELNRLVSQTKTDLCELYRETMLN----- 1253  
Db 1169 INQGLLEQDINTGKQTCYEBHGSYTPPLAVIKQSGF-----HYWHHCDSINAPL 1224  
QY 1254 EVRNGEMILLRTGTETIAQ-----RASKVLLTGTDSQOSVILTSDKQNLQSEAYSA 1306  
Db 1225 EVTNAQ-----GNTIWSGKYERFGRVSSPLSPYSSPER-----VMASFEQNL-- 1268  
QY 1307 YGKHSTANDASILGVNGRADPVSGVTHLGVNSYDPTLMRHTPDSLSPPFC--AGGIN 1365  
Db 1269 -----YAGQYFNETGL--HF--NTFRYDQIGRITMPD-----PIGLLGIN 1308  
QY 1366 PYSYCLGDPINRSDPSG-----HLNQWATGI 1392  
Db 1309 LYQYA--PNPLGWIDPWLQSSQEMVRVHRTSVEGLEGI 1345

## RESULT 7

US-09-252-991A-23805  
; Sequence 23805, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196, 136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23805  
; LENGTH: 1626  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-23805

Query Match 3.3%; Score 292; DB 4; Length 1626;  
Best Local Similarity 21.1%; Pred No. 4.6e-13;  
Matches 204; Conservative 125; Mismatches 316; Indels 320; Gaps 46;

QY 597 HRSQVRSY---TQRLPSDSTDITQIVTQMSYDKLGRLLTRTLNSGTPY----- 641  
Db 633 HTQVLAARYCNERGQLIEAGNALYAEERYDARNVILERQLAGGASFFWEWEGKQA 692  
QY 642 -----ANTLTVD--YELNN-----LQDNRPPFVITTTDVANGNOLRNE 677  
Db 693 RAVHWASFPQWDSHYVWNEGDSVTAINAGSEEVYHDDNAR--LVQVDPDGGETLH 750  
QY 678 FDGAGRHVSQCLKSDGKGFYTHHQYDQGR-----HHTSYSS----- 718  
Db 751 YDEKQLVAE--RDPLG---TITRYDQAGLEALPAEGEPTCYSYFDGFRVSRVR 803  
QY 719 -----DYLTNGR---QOTDPD--KVLHLSMSKSYDNGQIANTHWSYGVSEKITVDPITL 767  
Db 804 GEAQWKYERNAQGITROTDEGNVTHY---AYDRGCLVEITHADGSLHQLTWPLGQ 859  
QY 768 TATKQLOSNNNVQKGVTTTYSQPIQITLFEAGHLQSCHTLTRDGDWRVRKETDA 827  
Db 860 LIEQLPDGS-----VRRYRDTLGRQITRODESGAI----- 891  
QY 828 IQOCTIYQYDNNRVIOITLPGDIIVNRKYAPSTDTLITDIRVNGISLGQOTFDGSLRL 887  
Db 892 ---TRFQWDAAGRLSQITLPG--ASRYM-----RYN--AYGKVT----- 924  
QY 888 TQSDQGRVWATYTSAGNDQCPSTVITPDQOFIHYQYQPELDDAVLQVASNEITQOFSYN 947  
Db 925 SECDCQGRMTYEV--LDDLHLVSRINPDPGSQLRYR--ENARLLLSIENERGERYLD 981  
QY 948 PVTGALLKAVAGOSLTPYVPSGRLKMNINDMKMSYLVTLRG--LE-----NGVTDL 1000  
Db 982 -----YHNGLSIQCETGFCGRTAYRYDLKQLEKTEYDGDGSEL 1022  
QY 1001 TGTQKISR-----DTHGRVTOIKDSIKITLNYDVLNRRHIGSQVTD 1042  
Db 1023 RTTYQDSTGRLAKTLDPGNNRVYRDTLGRVAVDDGTWPLAYEYDLRDR----- 1074  
QY 1043 LATGHMLTTVE--FDGLNREIGRKLCD-----SSGHTL---DIQOSWLKTQOLANR 1089  
Db 1075 LVREHQWATLHYAYDALGQLIHCRLPDGNRVYRYQTGTLASIDLNGSLTRHQFGSG 1134  
QY 1090 IVKLVGLVQRTQYQYSDSRNLNOYKCDGABCTDKYGHISIVTQNTYDIYGNITACHTT 1149  
Db 1135 RERQOQCELLSQYHYDEQGLLAHQVSQKQ-----RHLYQRYRYDASGNLAA----- 1183  
QY 1150 FADGTEDHATKFAFNPDPCCOLTEVHHTHPDMPDNLRYDKAGRVIN----- 1197  
Db 1184 IEDSRKGIERSPHY--DPLD--RLJGVGRGETPE---SFVHDPAGNLLAQGGQFARQMEV 1235  
QY 1198 -----IT-----DNHGN-----TENTYDTLGR-----LQNGQSVYGYD 1227  
Db 1236 RGNLLTQGBRHFYDAHGNLVRERRGTGOKLVTEYSYDQOHLIGVSPDGRQVYRYD 1295  
QY 1228 PLNRLVSQKTDLCELYRETMLNVRNGEMIRLRTGTETIAQ--ORASKVLLTGTD 1285  
Db 1296 AFGRIAKQVDG-----RNTEFLWL---GERLLAESGRHRYKTYLYEED 1336  
QY 1286 SQOSVILTS-----DKQNLQSE-----AYSAYGK--HKSTANDA 1317  
Db 1337 SFRPLALDGEQPEQVPCYQDHLGTPQELTRAGRLCWSARYRAYGNVLKLDIAEVD 1396  
QY 1318 SILYNGERADPVSGVTHLGVNSYDPTLMRHTPDSLSPPFCAGGINPYSYCLGDPINR 1377



Db 1397 NPLRFQGGYDAETGLHY--NHRYYNFSTGRYLTPDPRL--AGGLNSYRY-VENPTGW 1451  
QY 1378 SDPSG 1382  
Db 1452 VDPLG 1456

RESULT 8  
US-09-328-352-5542  
; Sequence 5542, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Berton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-032A  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5542  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5542

Query Match 3.2%; Score 285.5; DB 4; Length 1596;  
Best Local Similarity 18.1%; Pred. No. 1.4e-12;  
Matches 202; Conservative 179; Mismatches 387; Indels 349; Gaps 49;

QY 440 ESNGLPKVKKDKTKIISPTWETYPAGEVDNCPPEPGYTRFVKKIOTPYDSEF 499  
Db 337 QSGKSIYSIGAERVQ-----HADFYLP-----KIGFS-PIRQ-----YNSQM 373  
QY 500 KDDPEKTIQRYSL-----IGSQSHVTLKIEERHVSATOLLNSTLFOYNTDKSELGRL 552  
Db 374 DEFQSVGARGWMPFNMQQNAGYLFDSKGRKH---QLPVSIIFFEYVEPYE-GWI 429  
QY 553 LK-----QTECTKGNGKTYSVVHKFTYTKQDDTLQOSHSHITTH-----591  
Db 430 IKPLKNGELILDFGGEWRSHFQSPDGKGYLVKMMNETSQEBELLE-YLLDHLIAVLKV 488  
QY 592 DNFTIHRSVRSRYTGRLFSDTDKDIVTQMSYDKLGRLLTRINSPTVANTLTVDYEL 651  
Db 489 INFKLQAEVELKP-----AFNEQVKIIAVFL-----515  
QY 652 NNLODDNRPFVITTTDVNGNQLNEPFGAGRHVSQCLKSDGDKGFYTHITQYDEQGR 711  
Db 516 ----DDKAEPLA-----RYEYDTQG-----NLKALDQNG-----HRTYENQF 551  
QY 712 HHTSTYSYDLTNGRQTDPPKVLHSMKSYDNGQIANTHWSYGVSEKITVDPIITATK 771  
Db 552 HQUTRYTD--RTGRGQNIYESTEAKAIAEAWDDGSEH-----TKLKWHP-----RLR 599  
QY 772 QLOQSNMNVQTKGVITYPESQPIQITLFDAGHLQSCHT-----LTRGDWRVKE 824  
Db 600 QVA-----VYDAVDVPT-----YFDDLGFTYRNLADGRSWTSRQKGRITQ 645  
QY 825 TDAIGQCTIYQDNNRNVIGITLPGCTIVNRKYAPFTDTLITDIRVNGISLQGFQDGL 884  
Db 646 IDFDGRETOQVNDQDLVKIVQPNGGIIRFAY-----NQK 681  
QY 885 SRLTQSDQ--GGRVWATYSAGNDCPSTVITPDGQFIHYQYQ-----PELDDA---VLQV 935  
Db 682 GNLVEIKDPGSLWKREYDENRN--VSKEINPUGHITQYKNNNDNOLVEIDAKGVKKI 739  
QY 936 ASNEITQFQSYNPVTGALLKAVAEGQSLPIYFSGRLKVENINDMKMSYLVLTWLRGLEN 995  
Db 740 QYNELQMISYTD-----CSGKSTWEYDEDCALTAEQTANNKVQYFYSYKGRDK 790  
QY 996 GYT-----DLGTIQKISRTHGRVTQIKDSSIKTTLNVDLNRH-----1035  
Db 791 GLOQSIYDPGLKEYFEHDEGRLLK-HTDTKGLVTEYKYNQVGLLEQRIDANRHSVAYQ 849

QY 1036 -----IGSQVTDLATGH-----MLTTTVEFDGLNREIGRKLCDDSSGHTLIDIOQSWL 1081  
Db 850 WDKQGRIOKLINQQAELFGYNPYGYLIREQAFDGBEKHYS---YNGRLEFQIRRPNI 906  
QY 1082 KTO-----QLANR-IVKLVGLQTEQSYSDSRNRLNQYKCDGACBPTDKYGHSI---1130  
Db 907 LTOFDYYADGQIASKSFTHLTGQKOTEQFDYNLNSQLSRASNEVSEQ--IDLRYNALGQL 964  
QY 1131 -----VTQNFETYDIYGNI-----TACHTTPADG-----TEDH 1157  
Db 965 VREHQHYKPELKPALTAVLHVEYDELGNLIKIRPDCGHTLNHLVYGGSHIYAIGLNNQEV 1024  
QY 1158 ATF-----KFAIPTDFCQLTEVHTHTPMDPNIR 1186  
Db 1025 VSPQDDLHRETRILLANGLMOTKQYNDVGLLSQFQEQETQDVLQYQAHN-----K 1078  
QY 1187 LKYDKAGRVINITDNHNTENTFYDTLGRQLONGQ-----SVYGYDPLNRLVQKTDLTLC 1242  
Db 1079 YHYDKNYLLSQVEDSRGLKLYQYDPIGRLLIAAQSLLHKTESFNPAGNLIDSESVLSPA 1138  
QY 1243 ELYRETMLVNEVRNGEMIRLLRTGETIIAQORASKVLLTGTDSQSVILTSDKQNLQOE 1302  
Db 1139 QI-----KNNLIKSYKGKHQYQYDQGNVTEIIQ--AGKNLKLTDWNCNRLIRSDNGLVTEY 1193  
QY 1303 AYSAYCK--HKSTANDASILYNGERA-----DPVSGVTHLNGNGYRS 1342  
Db 1194 GYDVFGRRUYKKTAKELTLFGWDGDLMIWESFKSAQTNYTKHYIYEPDSFVPLLOAGYKD 1253  
QY 1343 YDPTLMRHTPDSLSPFGAGGINPYSYCLGDPI-NRS 1378  
Db 1254 F---IQLIETPD-----YQEQYTKPYS-IVKDPVWNRN 1282

RESULT 9  
US-09-252-991A-23640  
; Sequence 23640, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23640  
; LENGTH: 974  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23640

Query Match 3.1%; Score 277.5; DB 4; Length 974;  
Best Local Similarity 20.4%; Pred. No. 2.5e-12;  
Matches 218; Conservative 141; Mismatches 377; Indels 335; Gaps 48;

QY 504 EKFTQYRYSLIGSQSHVTLKIEERHVSATQLLNSLTFOYNTDKSELGRLLKQTECTKGEN 563  
Db 112 ERWYSYNALG-----LIKRADGPRTDVQDVTLYAYD-----SRGNL 149  
QY 564 GKTYSVVKHFTYTKQDDTLQOSHSHITTHDNFTIHRSQVRSRYTG--RLFSDTDTKDIVTQ 621  
Db 150 TQVTNALGQVTRLDGYDERGKPGSITDANGVTSLSA-----YTGDGWLASVSTAGTTR 204  
QY 622 MSYDKLGRLLTRINSPTVANTLTLYELNLDNPPFVITTTDVNGNQLNEFDGA 681  
Db 205 FDYDAVGQITRVTRGDG---SWLSYEY-----DDARR--LVAIGNLGERLBYDVTDK 252  
QY 682 GRHVSQCLKSDGDKGFYTHITQO--YDEQGR-----HHTSTYS-DYLTNGRQOTDP 730

Db 253 GNTAQRKIDASG-----LVROQWAYDELGRLLRAVAGGQTRSFAYDLNDNPVGETNP 308  
Qy 731 DKVHLSKSYDNWGOIANTHWSYGVSEKITVDPITLTATKLOSNNSNVQTKGVITYT 790  
Db 309 RQ--FAHQAFDALDLVQSDPLGKTRLAYD-----AQNLTEVXDPKRGVTTRY 357  
Qy 791 PSQOPQITLFDRAHGL-----QSCHTLTDGWDVRKSTDAIGQCTIYCYDNRVNI 843  
Db 358 E-----YDGLGNLRLVSPDSGTTTFEHAAGNVIRRTDARGAVTEYRYDALNRLV 408  
Qy 844 QITLPGDTIVNRKYAFSTDTLITDIRVNGISLGGQTFDGLSRLTQSDGGRVWAYIYSA 903  
Db 409 ERSPSDPSLDVGR-----YDUTADGNK-----GRLG-GLDGHAR----- 444  
Qy 904 GNDQCFSTVITPDGQFIHYQYQPELDLDAVLQVNASNEITQOFSYNPVTGALLKAVAEGQSL 963  
Db 445 -----DSLVRYPD-ERGNLVEQVRSIRLDQQLDRLVTRYDAA-----NQL 485  
Qy 964 TPIYPSGRKLMENINDMKVSYLWTLRGLNGYDGLTGTQIKISRDTHGVTQIKDSSI 1023  
Db 486 LEIGYES-----GLAIGY-----PRNAGGQVAVS----- 509  
Qy 1024 KTLTNYDDLNHRIGSQVTDLATCHMLTITVPEFDGLNREIGRKLCDSSGHTLIDIOQSLKT 1083  
Db 510 -TLAVGDKAPSTLVGQIAYLPFGPLRLT-----WNGITLSREYDQD 551  
Qy 1084 QOLANRIV-----KLVGLOR-----TEQSYSDSRNLNOYK-CDGASCPDXY 1126  
Db 552 YQLRQKPGWSDYQHDANGNIQHRSLWGLTDYQDPLDRUTEERGVOG----- 604  
Qy 1127 GHSIVTQNTFYDIYGNIT-ACHTTTFADGTEDHATFKFANPTDPCQLTEVHHTHPMDPNI 1185  
Db 605 -----RSVAYDAVGNRTQSDNPASGCTASSQDYQA-----PDSN 640  
Qy 1186 RL-----KYDKAGRVNITDNEHNTENFTYDGLGLOL-----GQGSYVGYDPL-NR 1231  
Db 641 RLTAIGAQAQVTSRAG-----NLQDRA-ARKLAYDAQRLQSLDGGQVQAVRYNALGQR 696  
Qy 1232 LVSQKTDITLCELRYRETNMLNVEVRNEMIRLRTG-----ETIAQAQPAS----- 1277  
Db 697 IVKLTPESVTLYLPGDQLGAEHSGRKLRAQYLLWLDLPLATIDADYDAQKVG 756  
Qy 1278 --KVALTGTDSQSVLTSDX--QNLQSOEAYSAGKHSTANDAS--ILGYNGERADPVS 1331  
Db 757 NPTLLVHGHDITPRLATDASQIAWQSDAFERGEALSGGSTQVNLRPFGQYVDAES 816  
Qy 1332 GVTHLNGVRSYDPP-----TLMRFTPD-----SLSPFGAGGINP-- 1366  
Db 817 GLRY--NYFRDYPETGRYVESDPVETLRKLNPNPMTFLNGESMLQATPYWEHGFTEPH 874  
Qy 1367 -YSYCLGDPINSD-----PSGHSW-----QAMTIGMGFIAGLLLIAT 1405  
Db 875 NYTSDNNFTAKSDKHGSLSPNTDNLITPDNTCTCTLECKKFTGKXSF----- 925  
Qy 1406 GGVATAAAGGIAAATASTTALAFGALSVTSDITSIVSGALEDASPASS 1456  
Db 926 -----LVGALCKATTFE-FGVVVCNSTIWMICGASCQSCQECNRAPS 965

## RESULT 10

US-09-543-681A-4476  
; Sequence 4476, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILII  
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709,1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4476  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4476  
  
Query Match 3.1%; Score 271.5; DB 4; Length 1665;  
Best Local Similarity 20.7%; Pred. No. 1.9e-11;  
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;  
  
Qy 508 OYRSLIGSQSHVTLKIBERHYSATQNLNSTLFOYNTKSELGLLKT-----EC 558  
Db 608 QIRIDFIINQSOLVSV-KRHDGIALIL-----EYQDK-LVRICSESTSPRPFVRC 659  
Qy 559 TKGNGKTYSVVHKFTYTKQDPTLQOSHSHITHDNFTIHSQVRSRYTGR-----LFSDDT 614  
Db 660 EYDTHG-----YLSQCHAYQNH-----LWHRYSPEGLMVAWGDTD 695  
Qy 615 TKDIVTQMSYDKLGLLTLRTLSG-----TPYANTLTY-----DYELNNLQDNRPP 661  
Db 696 STELT--IDYDEQKRVVATHSPSGFWDRIYDDYQRMITYIDABGGSFYVYNDN-- 750  
Qy 662 FVITTT-----DVNGNQLRNEFDGAGRHSQCCLKD---SDGDKFY 699  
Db 751 -LVTRTIDPLWRETVTWEQKRIAEINEIGERTEY---GYHNGULLAVIYLPDKAIYV 806  
Qy 700 TIHTQOYDEQGR--HHTSTYSYDYLTLNGRQOTDPPKHLSMSKSYDNWQIANTHWSYGV 757  
Db 807 -----DYNDYQOLTHFTSAFGD-----EWOLSYDENGNT-- 836  
Qy 758 EKIVTDPITLTATKLOSNNSNVQTKGVTT-----YTPSQPIQITLFDGAGHLOS 809  
Db 837 -LVTRP---QGRQVVEYSQSGELLKAITPNGAQWQYHYNPAHLIKTT--NPYQHS 889  
Qy 810 CHTLTRCGWDRVRETDAIGQCTIYQ-----YDNNRVNI 843  
Db 890 YHS---DELGLLHYTALNHTTRYQYSTEASTNGSLSKILLPDGVEQIDYDSERV 946  
Qy 844 QITLPGDTIVNRKYAFSTDTLITDIRVNGISLGGQTFDGLSRLTQSD--GGRVWAYT- 901  
Db 947 AVTDCEGKTRTRYRYPGP--DVLLAMIRPDGSEI--RPEYDSLRLKLVVANGSEVLYER 1003  
Qy 902 SAGN-----DOCESTVIT--PGQRTHYQYQPELDAVLQVNAS- 939  
Db 1004 KAGQIIRVDFTGREICVRYDRLGRRIATRYPDNHELWRVY---NESGLVVEQSEW 1059  
Qy 940 -----ITQOFSYNPVTGALLKAVASGQSLTPIYPSGELKXENINDMKMSYLTLRG 992  
Db 1060 EQESRCLSTAQSYN-ARQQLIKATNPDSVVEFEYDDQGLCSERINE-QEIVHOW- 1113  
Qy 993 LENGVTDLTGTIKI-SRDTH-----GRVTQIKDSSIKTTLNVDLNRHIGSQVTDLAT 1045  
Db 1114 ---NEADNTLALTFRGRELHYAFGALGELTSQ-----VNQHAPQFS--- 1154  
Qy 1046 GEMLTITVFEGLNREIGRK-----LCSGHT-----LDIQOSWLTQQLANRIVKLVGL 1097  
Db 1155 -----YNAVQCEYLRRSRAGFVNSSHYTATGLAHQORAGRTEQF-----LQSIQ 1199  
Qy 1098 RTEQ-----YSYDSRNLNOYKCDGACPTDKYGHSTVTONFTYDIYGNITACH 1147  
Db 1200 AHPQOPPECTDVHRSYQYD-----RAYVVGIE--DDRWRQT---RYHYNANDQITETQ 1248  
Qy 1148 TTFADGTEDHATFKFANPTDPCQLTEVHHTHPD--MPDNIR-----LKYKAGRVINI 1198  
Db 1249 YSPQWGNODE-KFOYDNNLN---ITE-HLTPSSSMVPSDAQAMLQFQQQAGSVTR 1303  
Qy 1199 TDNHNENTFTYDTLGLQ-----NGQSV-----YGYDP 1228  
Db 1304 YTAGK-YQDYHYDNGRLAKKIIVHTRGFRPREWRYLWNTQNLCTACFTPKGDCWHYTYA 1362  
Qy 1229 LNRLVSQKTDLCELRYRETNMLNVEVRNEMIRLRTGETII-----AQ- 1274

Db 1363 FGRLS-KTKVDSLALHIDPLFPQIKITWYLVSGDLIEETPIYADGTLANAQV 1421  
QY 1275 ----RASKVLLTGTDSQSW-ILTSKQNLKQEAAYSAYG-----KH 1310  
Db 1422 QWLYQGEHTPTARYQCKLHVYVTHQGTPTREIFSEGQASWAGRLNTWGMQFWRYRD 1481  
QY 1311 KSTANDASI-----LVNGERADPVSGVTHLGNYSYDPTLMRFTPDLSPPG-AGGIN 1365  
Db 1482 GRAENDPNTECFPFAGQVEDEESGLYY--NRFYRYDRETQYLSFD---PIGLLGLN 1536  
QY 1366 PYSY--CLGDPINRSPSGHLSQAWTGICMGIAG-----LILTIAT 1405  
Db 1537 PYGVYHC---PGWVDP-----FGLAGDCDELLVLAT 1566

RESULT 11

US-08-851-567B-61  
Sequence 61, Application US/08851567B  
Patent No. 6528484

GENERAL INFORMATION:

APPLICANT: Ensign, Jerald C  
APPLICANT: Bowen, David J  
APPLICANT: Petell, James  
APPLICANT: Fatig, Raymond  
APPLICANT: Schoonover, Sue  
APPLICANT: French-Constant, Richard  
APPLICANT: Rocheleau, Thomas A.  
APPLICANT: Blackburn, Michael B.  
APPLICANT: Hey, Timothy D.  
APPLICANT: Merlo, Donald J.  
APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Ciche, Todd A.

TITLE OF INVENTION: Insecticidal, Kitistri  
NUMBER OF SEQUENCES: 88

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seav, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.93804

TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1043 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-61

Query Match

Best Local Similarity 20.4%; Pred. No. 3.1e-09;  
Matches 221; Conservative 130; Mismatches 378; Indels 353; Gaps 51;

QY 612 DTDKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTVD---YELNNLQDDNRPPFVIITTD 668  
Db 38 DTDTR--VTRHYDARGHL-----NVSIDPRLYDAKQADNSVKPNFV-WQHD 81  
QY 669 VNGNLRNEFDGAGRHYVSCCLKDSG-----DGFYTIHTQY 706  
Db 82 LAGHALATESVDAGRTVA--LNDIEGRSVMTMATGVQRTRRYEGNTLPGRLLSVEQVF 139  
QY 707 DEQGRH-----HTSYSDYLTNGR--QOTDPKVKHLSMSKSY-----741  
Db 140 NOESAKVTERFIWAGNTTSEKYNLSGLCIRHYDTAGVTFLMSQSLAGAMLSQSHQLAE 199  
QY 742 ---DNMGQIANTHWSYGVSEKIIVDTITLTKQLQSNVNVOTGKEVITYTPSQOPIQI 798  
Db 200 GQEANWSGDDETVMQGLASEVY-----TOSTTNAIGA---LLTQTDAGNIQR 246  
QY 799 TLFDEAGHLQSCHTLTRDGMWRVRKETD-----AIGQCTIYQYDYNRVIQITLP 848  
Db 247 LAYDIAGQL-----KGSWLTVKGQSEQVIVKLSWSAAGHLKREHGN-----289  
QY 849 DGIIVNRKAPSTDTLITDIRVNGISLQQITFDGLSRLTQSDQGRVWAYTYSAGNDQC 908  
Db 290 -GVVTEYSYEP-ETQSLI-----GIT-----TTRAEGSQSGARVL-----322  
QY 909 PSTVITPDGQFIHYQOPELDDAVLOVASNEITQOFSYNPVTGALLKAVAEQ-----SLT 964  
Db 323 -----QDLRYKYDPV--GNVISIHNDAAEATFRNQ-----KVEPNRYVYDSLY 365  
QY 965 PIYPSGRKLMENNDMKM--SYLWTLRGLNGYTDLTGTIKISRDRTHGRVTKDSS 1022  
Db 366 QLMSATGR-EMANIGQSNQLPSPVPTDSDSTVNY--LRTVYDRGGNVLQIRHSS 421  
QY 1023 IKTLNY-DLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLIDQQSWL 1081  
Db 422 PATQNSYTTDIT--VSSRSNRVAVLSTLTPTDPRVDA-----LFDSCGH-----QKMLI 467  
QY 1082 KTQQLANRIVKLVNGVLRQTEQYSYD--SRNRLNQYKCDG---AECPTDKYHSIVTQNP 1136  
Db 468 PQNLWDNI---RGELOQVTPVSRNSSDSEWYRSSDGMRLLYKVEQQTNGSTQVQV 524  
QY 1137 YDIYGNITACHTTFADGTEDHATFPANPTPCQUTTEVHHHPDMNIRKYKAGRVI 1196  
Db 525 Y-----LPGLELRTGVADKTLEDLOVITVGEAG---RAQVRVLHWSGKPT 568  
QY 1197 NITDNHGNTEFTYDTLGRQNGQSVGYDPLNRLVSOKTDTLDCELYRYETMLVNEVR 1256  
Db 569 DIDNNQVR-----YSYD--NLLGSSQLE-----589  
QY 1257 NGEMIRLLRTGETTIIAQORASKVLLTGTDSQSVILTSKQNLKQEAAYSAYG-----KH 1310  
Db 590 -----LDSEGOILSQEYYPYGGTALWAAR 614  
QY 1311 KSTANDASILGVNGERADPVSVTHLGNYSYDPTLMRFTPDLSPPG-AGGINPYSY 1369  
Db 615 NQTEASYKFIYSGKERD-ATGLYYY--GYRYQFWGRWLSAD---PAGTVDGLNLYRM 668  
QY 1370 CLGDPINRSDPSG-----HLSWQAWTGIGMGIAGLLTIATG---GMAIA-----AAG 1414

Db 569 VRNPITLTDHGLASPNRNRNFWFASFLFRKPDGMSASMRGQKIGRAIAGGIAIG 728  
 Qy 1415 GIAAIASTSTAL--AFGALSVTSDITSIVSGALEDASPKASSILGWVWGM-----1465  
 Db 729 GLAATIAATAGAAIPVLGVAAVAGAGALMGVYVSGLLERGAALLARLVQOKSTLVQSA 788  
 Qy 1466 --GAAGLAESAIG--GFKIATHLGAFAEDGENALLKSTSESSRIKWGVTRSLDREIV 1519  
 Db 789 AGAAGASSAAAGAAAGVGVASAAAGVT-----GAVGSWINNAD 829  
 Qy 1520 RNEGVQIKHDSRYTNDWNGEQAIIWH-----GDKOGLYHTEGK-----HNGKG 1568  
 Db 830 RGIGGAIGAGSAGVTITDMLGTA--STLTHEVGAAGGAAGMITGTQGSTRAIGHAG 887  
 Qy 1569 PY 1570  
 Db 888 TY 889

RESULT 12  
 US-09-328-352-5545  
 ; Sequence 5545, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 5545  
 ; LENGTH: 804  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-5545

Query Match 2.7%; Score 235.5; DB 4; Length 804;  
 Best Local Similarity 20.8%; Pred. No. 3.3e-09;  
 Matches 147; Conservative 92; Mismatches 253; Indels 215; Gaps 30;

Qy 699 YTHIQYDEQBBHTSTYSDLTNGRCQTDKVLHLSMSKSYDNWGOIANTHS-----753  
 Db 21 FTLQTEYDAQAIRSNTQAGRVKYKYEY-----VSFSTCL---GNLASFQWSKVSC 71  
 Qy 754 ---YGVSEKITVDPIITLTKQLQSNNNVQTGKVTYTPSQ--PIQITLDFEAGHLQ 808  
 Db 72 LSNVVLGDQHVGLPIEV-----DVNGKEITQFAGYKRGIPQVKLANGA---116  
 Qy 809 SCHLTRDGDWRKRTDAIGQCTIYOVDNRYNVIQITLPGTIWVRKYAPFSTOTLTD 868  
 Db 117 -TETNIVDDFGNITQHTDAGVSRKQYDDAGRL---YIDPIVGLNYSTFTYDGTJVS 171  
 Qy 869 IRVNG-----IS---LQQQTFD-----GLSRLTQS- 890  
 Db 172 RVVTGGQLSRTEKNGDGLLSSBCKISNKSIIINSKYDAFGNLIKSNPCGSAITSGT 231  
 Qy 891 -----QDGRVWATYTSAGNDQCPSTVITPD-----GQPIH--- 921  
 Db 232 TSSYVDVDRPITVNDNGSVVTCYSCGGKTAIVQITDSTGTTESNLLAAGDFSALKT 291  
 Qy 922 -----YQYPELDDAVLQ---VASNEITQQFSY-----NPVTGALLKA 956  
 Db 292 LVARKGTGDSVFQTTTFEFENALLKPKVAVSGSSTQSYTNGNTTILATEKONSISGQTFK 351  
 Qy 957 VARGQSILTPYPSGRKMNINDMKMSYLWTLRGLNGYT-DLTGTQIKISR-----1009  
 Db 352 YDGTGRTSITPDSQSVETIKFQKDLASRTWREVTTSYSLAGRLKTTTIANISEA 411  
 Qy 1010 ---DTHGRV-----TQIKDSSIKTTINDDLNHRHIGSQVTDL-----A 1044  
 Db 412 FDLDTYGRVISHTQKINANDTNNVYVGYNQLN-----QVTSIQYENGKSVNLNQNA 466

Qy 1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSWLKTCQLANRIVKLVGLQR---TE 1101  
 Db 467 LGEVTSIPNVIOSLNRYARQQLTTVQANTDTLWSYTNDSGLLNNISATS--ERKCVLVN 524  
 Qy 1102 QYSYSDRRNLNOYKCDGACPTDKYHGSIVTQNTFYDIYGNITACHTTFADGTE-DHATF 1160  
 Db 525 DYGYDKLNRVNL-----SDKCGS---VYNATIDRYG-----TCLMSTVELDQARY 567  
 Qy 1161 KEANPTDPCQTEVHHTHPMDPNIRLK-----YDKAGRVINITDNGHNTENFTY 1210  
 Db 568 QYGYND--DIKV-----NITSSTVAPAIYTYNYVNTSRLASVSGSTISFTY 616  
 Qy 1211 DTLGRLONGQSGVGYDPLNRLVSKQTDITLDCELYRETMLVNEVRN 1257  
 Db 617 DAMGNVNDGVRTLTYDYSRLNKNENETY---LYNADGLRVRARD 660

RESULT 13  
 US-09-252-991A-31032  
 ; Sequence 31032, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 31032  
 ; LENGTH: 1128  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-31032

Query Match 2.5%; Score 223; DB 4; Length 1128;  
 Best Local Similarity 19.6%; Pred. No. 5.6e-08;  
 Matches 258; Conservative 151; Mismatches 467; Indels 442; Gaps 68;

Qy 235 NPSLGNENPLTWSFGYTP-----IGKNGILGOWITSMTPAGLKETVYNNNOGHH--PP 288  
 Db 63 DPALAAPDTFTFARGYLLSNPRIGRLG-RGWLPFGESMHLELSEDACVLVDAGRRIGFP 121  
 Qy 289 QSANLPVLPVYVTLMKQVPGAGQPAIQAEYSYTSYHNYVGGSGNGIWNKLNLYGLMTEYN 348  
 Db 122 ALA-----PGA-----QHYS-----GSEELM-----137  
 Qy 349 YGSTERRYKKEGHDQIVRIERTYNNVHLLTSECKQONGYIQTETAYVAITGHNFDSQ 408  
 Db 138 -----LRKGGSGGGEAQAWR-----GRWAAVPAB-----LOTQEGSVLVLSGHSY-- 177  
 Qy 409 PSQFOLPKTKTETWRSADNSYRSBIT-----ETTFDESGNPLTKVTKKK-----TQ 455  
 Db 178 -LHFQ-----RCPDGIWRLOASFGRAGYRTFWRSGRGLLTGVRDSAGRSYALVYQ 227  
 Qy 456 KLIISSTHWEYPPAGEVDNCPPEYGFTRFKKIIQTPYDSBPKO--DPEK-----FIQ 508  
 Db 228 QACEES-----EGDD-----GLRUF--GVILLASHDGPFPYIDPQSGLDWLVR 269  
 Qy 509 VRYSLIGSQSHVTLKIBERHYSATQLL-----NSTLFOYNTD-KSELGRLLK 554  
 Db 270 YQFSDSGD---LIAVRDLQVVRVFAWREHMLVAHCEPGGLEVRVYWDVHAPHGRVVK 325  
 Qy 555 QTECTKGNGKTVSVVHKFTYTKQDDTLQOSHSTTHDNTIHRSQVRSYTCRLFSDTD 614  
 Db 326 QIEAG-----GLTRTFYVLR--DATEVSDSLGRVERFEFAGEGGRWALVADGS 375  
 Qy 615 TKDIYQMSYDKLGRLLTRTLNSGTVPANTLTVDYELNNIQQDNRPFPVITTTDVGNNQL 674

Db 376 R-----SEFDYDLFGRL-----VAMRDPLGRET 398  
QY 675 RNEFDGAGRHVSQCLKSDGDKFYTHIQYDEQGRHHTSYSDYLTNGRQOTDPDK-- 732  
Db 399 RRRRDGQGRMLEE-----ESFGKARY-----EKYVDEETGL 429  
QY 733 -VHLSMSKSYDNWGOIANHWSYGVSEKITVDPITLTATKOLQSNNVQ--TGKEVTY 789  
Db 430 LVELE-----DAMORRTFDER-----GNATVIRGFASTRAY 465  
QY 790 TFSQOPIQIT-LFDAGHLOSCHTLTRGWR---VRKETDAIGCTTIQYDNNRVIQI 845  
Db 466 EDPRLPDAPTRIVDPGRGER-----RLEWRNFGLLAALATDCSGQWRYDYDNEGRIVAS 519  
QY 846 TLPDGTIVNRKYAPSTDTLITDIRVNGISLQGTFFDGLSRLTOSQDG-----G 894  
Db 520 SDPLGQLTERRYDPLG--QLIGLELADGALSIE-YDALGROTRIADAEGHATLFSWGHG 576  
QY 895 RWAWTYAG-----NDQCPSTVITPDGQFIHYQYQPE-LDDAVLQVASNEITQOFSY 946  
Db 577 DLLARVSDAGGELSILHDEAGRLVALTNENGVOAQFRYDLRLVEETGDFGRQRYR 636  
QY 947 NPVTGALLKAVAGQSILPIYVPSGRKVENINDMKM-----SYLWTLRG----- 992  
Db 637 NAADELIAREDAGRETTYADROGLASIRVPATEHAPALVERVRLADGLASAGAD 696  
QY 993 LENGVT-DLTGTQIKSRDTH--GRVTQIKDS-----SIKTTLYNDLNRHIGSOVTDL-- 1043  
Db 697 CEVRYTYDEVGNL-RLESOVHADGWYSVEHSHDALGVROTSRYGD-----APPVAMLY 750  
QY 1044 ATGHM---LTTTVER-----DGLNREIGRKLCDSSGHTLIDIOQWLKTOQ-----LAN 1088  
Db 751 GPHLHGALVGAVALAFERDALHREVRDA-----RRDQDDALFTQERQHAPLGLROR 804  
QY 1089 RIVKLVQIQRTEQSYDSRNL-----NOYKDCGACPTDKYGHSI-----V 1131  
Db 805 SRLRAGGDFWQGRYVDGLQLVIGIDNOY-----PSVRYEYDLCGLLASRRAGAA 857  
QY 1132 TONTYDIYGNITACHTTFADGTEDHATFKFANFTDPCQITVTHHTHPDNDIRLYKDYK 1191  
Db 858 ASTRYDAAGN-----RLEGVGEYAR-----EDARQAFANELYRSRFSRSEVRANQ 904  
QY 1192 AG-----RVINITDNHGNENTFYDTLGRIL-----ONGQSVYGVYDPLNRLYSQKT 1237  
Db 905 AGEQPARWAGNRVERIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAORLVHLTR 959  
QY 1238 DTLD---CELYRETMLNVEV-----RNG--EMIRLLRTGETIIAQORASKVLLTGTDSQ 1287  
Db 960 DYADGTREARYRYDALSRRIAKVLRDQVEQVRFQWGDGDRQCAEAFARELATTVHEPG 1019  
QY 1288 QSVIL-----TSDKQNLQO--EAYSAYGKHKSTANDASILGNGERADPVSQVTHLGN 1338  
Db 1020 GFVELLRLQACEPDPPELQLQRAFAEGQPL-----PAQCVPALGE 1062  
QY 1339 GYRSYDPTLMRFHTPDSLSPPFGAGGINPYSCYCLGDPINRSDPPSHLSWQA-----WTGI 1392  
Db 1063 ARIAF-----FHTDH-----LGTPLQLSDERGQLRWQGVDPDDWRAV 1098

## RESULT 14

US-09-328-352-5486  
; Sequence 5486, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Berton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5486  
; LENGTH: 3290

TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5486

Query Match 2.5%; Score 222.5; DB 4; Length 3290;  
Best Local Similarity 18.8%; Pred. No. 3.9e-07;  
Matches 370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;

QY 12 ITWSDNNEF-----FTQANNFTSAVG-----GVDPR 39  
Db 459 LTLSDLNDTGVSATDQITQDKNFNLKLEGQETGQVTVLVSTDEGKTWQETTVAOKDLAD 518  
QY 40 GLYNIQITLCHIVGNGN-----LGPTLPLT--LSYSPLNKTDIGFGIFGNFGLSVVD 89  
Db 519 GIYQYKAVTDVAGNISETAIKQVVDVTPAQACKLTLSDLNDT-----GVSATD 568  
QY 90 ---RKNS-----LLSSTGE-----NYKVIETDKTVKLOQKLDNLRFFKDLKENCYRII 136  
Db 569 QITQDNSFTLKLQPIVIGEQAAALDHYEVSDEGKTWQETTAD---QKLDADGIYQ-- 622  
QY 137 HKSGDI:EVLTGFNNAFDLKVPKLLNPAGHAIYIDWNEATQPLNRIRYDLDGDHDIPL 196  
Db 623 YKA---VITDLAGNISESAIOKVVDNS-----LNVESTIVIVKPIED----- 663  
QY 197 LNLVYQGLIKTILTLFPQKQEGYRTELRFNRLQNSIHNFSNGENP-----LTWSEGYTP 252  
Db 664 ---NTISLVEKDQVISIRLBIAIANLPTDLNS---SLTSVNTLTLENVTYNPHFDE 710  
QY 253 ICKNGILGOWIT-----SMIAPGGLKETVYNNQGHHPQOSAN 292  
Db 711 VTQ-----EWTEIPAELFWSVEPQTNISIDISLTDAQNTAITHTQN---YNDPTFN 762  
QY 293 LPVLVVTLMKQVPGAPAIQAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNYGST 352  
Db 763 SPILDSLT--FNLDGA-----LISGAYKSKVDIYNKNGDWLASTIT----- 804  
QY 353 ESERYKDEGH-----DOIVRIERTYNNYHLLTSECKOQNGYI--QTTETAYAI-- 400  
Db 805 ---NEEGKFTLODLSINTNQEVAVATYNGY-----SSNSSIGLAVTEVPAISIT 851  
QY 401 ---IGHNFDQSPQFOLPKTKTETWRSADNSYSEITETTFDESG----- 442  
Db 852 RISPEGVISGYATESHFIV-----KDQNGNLQEFNSNVFDSGITPPFSVMALEGV 903  
QY 443 ---NPLKTVIKDKTKQKILSP-----STHWEY-----YPPA-----GEVD 474  
Db 904 RPFILSLDQPL-----EEGAQIIISTDKDNISCHQYITADYTPAVFLETQPDISGETL 958  
QY 475 NC-PPEPYGFTFRVK---KIIOTPYDSFEKDDPEKFIQYR-YSLIGSQSHVTLKIEERHY 529  
Db 959 SVHVNEPNSFIRAFSGEGNLIATGF-----TDEQGFASLQVFPQFLKGETVTVQVVDKNQ 1013  
QY 530 SATQLL-----NSTLFPQYNTDKSELGRLLKQTECTKGE 562  
Db 1014 NTSSETLIEVPNFAYIPHVERITQEGLSGVAEDNSTVIVROADGNELKV----- 1063  
QY 563 NGKTSVVKHFTYTKQDDTLQOSSHITHTDNFTHRSQVRSRYTG-----R 608  
Db 1064 ---TLGDD---NSWSDFGHFSLSVNRPLIDGEEKISVQIIDNKLMSPEQN 1107  
QY 609 LFSDDTDTKDIVTQMSYDKLGRLLTTLNSGTPYANTLYD-----YELNNLODD--- 657  
Db 1108 IIVDLTPPPAPTELNFNDAGDLV---YGHAEFFSEILVKDQGNILNKFNNWNTDESGS 1164  
QY 658 ---NRPPVITTTDVGNG-----QLRN-----EPDG-----AGRHV 685  
Db 1165 FSIELGTFLTNAETVYVVTATDVNGNVSLAAQIAPNAPAFYVDSFTSDGVISGOAENS 1224  
QY 686 SCLKXDSGDGKFTYTHIQYDEQGRHHTSYSDYLTNGRQOTDPDKVHLSMSKSYDNWG 745  
Db 1225 TLVWAKAGD---VVAEIKVGEENGWNGSSIFK-LQLDRPLVDGEQFFLSIK---DARG 1276  
QY 746 QIANTHWSYGVSEKITVDPITLT-ATKOLQSNNNVQTG-KEVTTYTPSQOPIQITLFEDE 803

Db 1277 QVS-----ADVTITADTAPASNLVSEDSYLTGVAELNT-----TIQVFDH 1321  
Qy 804 AGHLQSC--HTLRDGNDRVRKETDAIGCTIY-----QYDYNRVITQ 844  
Db 1322 NGQLVNIWNTNSD-----GTFIYLGNNLHGEAFTVTKDQAGNVSEALS 1369  
Qy 845 ITLPDGTIVARKAPFSTDLITDIRVNGISLGO-----TFDGLSELTSQSQGRV 896  
Db 1370 INAPLDDI-----APNPKILLD--ANGCNFTAQAANSQIEVFDSLQ-----NQTG--- 1415  
Qy 897 WAYTYSAGNOCQSTVITPQGFTH-YOYQPELDDAVLVASNE-----ITQOFSYN 947  
Db 1416 WGSTDSAGN-----ASGSFNQTYLHGEELTFVVIDRAGNRSIEFKQNALIDTIAPN 1466  
Qy 948 PV-----TCALLKAVAGQSULTPIYPSCRLKWMENNDKXSYLWTLRGLNGVYD 1000  
Db 1467 PIENIIFNENGQSTQAAGAGSSIDVLDQGT-----NKIGF-----GYTDS 1507  
Qy 1001 TGTQKISRDT--HG-----RVTOIKDSSIKTLLNYDDLNHRHIGSQVTDLAT 1045  
Db 1508 SGNVSGYFQYVYLHGEELTFVVIDRAGNRAEVKQSA-----LNDVAPNPIENILLDL-N 1562  
Qy 1046 CHMTJ-----TTFEPLGNRE-IGRLCDSSGHTLD-IQOSWLKTOOLANRIVKLGVL 1097  
Db 1563 GQNFATAQAANSQIEIKNNNGDVVGYGSADSNVSGYLYOVHLHGEELTFVVDLAG-- 1620  
Qy 1098 QRTQSYSDSRNRLNQY-----CDGAECPTDKYGHISVITQNF----- 1136  
Db 1621 -----NRSTEVKQNALIDDIAPNPIENIVLDINGQNFATAQAANQIEVKNVAG 1669  
Qy 1137 -----YDIYGNITA-CHTTFAGDGTEDHATFKFAN-----PTDPCQLTEVHTHDPMD 1183  
Db 1670 EIVSGYVDGAGNVSGYLYVYLHGEELTFVVIDRAGNRSIEVKQNALIDDIAPNPIE 1727  
Qy 1184 NIKLYDAGRIVINTDNGHNTNFT-----YDTLGRQNGQSGYGYDPLNRLVSKQTD 1239  
Db 1728 NILL-----DANGQNFATAQAANTQIEVKNVAGEVIG-----SGTDS 1765  
Qy 1240 L-DCELYRETMLVNEVANGEMIRLL-----RTGETIIAQARASKVLLTGTDSQSQSVLTSD 1295  
Db 1766 MGNVSGYFQYVYL-----HGEELTFVVIDRAGNRSIEVKQNALIDDIAPNAIENIIFEN 1820  
Qy 1296 KQNLQSEA-YSAKGHKSTANDASILVNGERADPVSG-----VTHLGN 1338  
Db 1821 GQNFATAQAANSQIEVKNVAGEVIGSY-VDSAGNVSYLNOVYLKGEELTFVVIDQAGN 1879  
Qy 1339 GYRSYDPTLMRFHTPDSLSFPGAGI-----NPYSYCLGDP-----INRSDPSGHL 1384  
Db 1880 --RSIEVVKQTAFLDNTAPENATNLVSEDSYLSGMAEPNATQIFDQVQGLLNOWN 1935  
Qy 1385 --SQAQWATIGMG-----IAGLLLTATGMAATAAGGIAAATASTTALAFGALSVTSDI 1439  
Db 1936 NYNWDGTFNLYNSNYMEGFVKV-----VVVDQAGNLSGEV----- 1973  
Qy 1440 TSIVSGALDEASPKASSILGWVSMGGAAGLAESAIGKGTKLATHLGAFAEDGSENALLKS 1499  
Db 1974 ---VKAPLDDIAPVAASDLVFNEDGSSLSGVAEP-----NTFIQIFDQNGQQ--MNT 2020  
Qy 1500 TSESSRIKWVTRSLDREIVNEBQVVKDHSRGVTDNFMKGKGCQAILVHGDKGFLYHT 1559  
Db 2021 WSQSVNADGTFITFFGTYNLHGEETFTI-----VKDLAGNVSEAVSV----- 2062  
Qy 1560 EGNKENGKGPYTRHTPEOLVDYLDKNNIVDLTQGGD--KPVHLLSCYKSSG 1609  
Db 2063 -----KAPLDDIAPNFIKNIVFDANGQSTQAQAANSQIEIFDSFGSQIG 2107

RESULT 15

US-08-447-031A-2

; Sequence 2, Application US/08447031A

; Patent No. 5851794

; GENERAL INFORMATION:

APPLICANT: GUSS, Bengt  
APPLICANT: HOOK, Magnus  
APPLICANT: JONSSON, Hans  
APPLICANT: LINDBERG, Martin  
APPLICANT: PATTI, Joseph  
APPLICANT: SIGNAS, Christer  
APPLICANT: SWITALSKI, Lech  
TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS  
TITLE OF INVENTION: ITS PREPARATION  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,031A  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/861,804  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/SE91/00707  
FILING DATE: 22-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9003374-7  
FILING DATE: 22-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 012889-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1183 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-031A-2

Query Match 2.5%; Score 220.5; DB 2; Length 1183;

Best Local Similarity 19.1%; Pred. No. 9.6e-08;

Matches 206; Conservative 159; Mismatches 405; Indels 311; Gaps 56;

Qy 299 VILMKQVPAGQ---PAIQAEYSYTSNHYVGGCSNGIWNKLDNLYGLMTEYNYGSESR 355

Db 205 ITIKDQIQGGQLDLSTLNINVTGTHSNYSQS-----AITDFEKAFFPSK 251

Qy 356 RYKDEGHQDIIVRIERTYNNYHLTSECK-----CQNGYIOTTETAYAIIG----- 402

Db 252 ITVDNKTNTIDVTIPQGYGYSNFSINVKTKITNEQKQEFVNNNQ-AMVQEHKEEVNGK 310

Qy 403 -----HNFDSQPSQFQLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVKKDKTKQ 456

Db 311 SFNHTVHNINANAGIEGTVKELKVLKQ-DKDTKAPIANVKEKSKD-GSVVKD--NQK 366

Qy 457 IISPSSTHWEYPPAGEVDNCPPEPYGTFRVKIIQTVPDSEFKDDPKFIQYRSL--I 514

Db 367 EI-----EIIIDANGIANIKALPSG--DYILKEIEAPRPYTFDKKE-----YPTMKDT 414

Qy 515 GSQSHVTL-----KIE-----ERHYSATOLLNSTLF-----QYNT---DKSELGR 551

Db 415 DNQGYFTTIENAKAIEKTKDVSQAQWEGTKVKFTIYFKLYKQDDNQNTTPVDRAEIKK 474







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:44:38 ; Search time 67 Seconds  
(without alignments)  
7772.803 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLRITMSDNNF.....PRKIILGRTEKTVKPKTRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCT05\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932.5	10.5	982	12	US-10-365-742-106
2	517.5	5.8	1317	15	US-10-369-493-11243
3	420.5	4.7	2364	14	US-10-156-761-7834
4	395	4.4	820	14	US-10-156-761-7990
5	392	4.4	1250	14	US-10-156-761-7752
6	391.5	4.4	2386	14	US-10-156-761-7751
7	386	4.3	2234	15	US-10-282-122A-46565
8	381	4.1	1510	15	US-10-418-861B-55
9	353.5	4.0	843	12	US-10-282-122A-76214
10	350	3.9	1515	12	US-10-282-122A-47600
11	342	3.9	1397	12	US-10-282-122A-43059
12	340.5	3.8	1411	12	US-10-282-122A-43060
13	339	3.8	1426	9	US-09-912-020-340
14	339	3.8	1426	12	US-10-282-122A-42617
15	331.5	3.7	1377	9	US-09-815-242-10384

16	331.5	3.7	1377	12	US-10-282-122A-42731	Sequence 42731, A
17	331.5	3.7	1377	14	US-10-287-274-467	Sequence 467, App
18	328	3.7	1572	12	US-10-282-122A-69415	Sequence 69415, A
19	322	3.6	1385	12	US-10-282-122A-68242	Sequence 68242, A
20	317.5	3.6	1565	12	US-10-282-122A-44558	Sequence 44558, A
21	314	3.5	1395	12	US-10-282-122A-50586	Sequence 50586, A
22	307.5	3.5	932	12	US-10-282-122A-44617	Sequence 44617, A
23	306.5	3.5	1530	12	US-10-282-122A-69175	Sequence 69175, A
24	298.5	3.4	1531	12	US-10-282-122A-49308	Sequence 49308, A
25	297	3.3	1512	12	US-10-282-122A-78386	Sequence 78386, A
26	291.5	3.3	1253	12	US-10-282-122A-68914	Sequence 68914, A
27	291	3.3	2346	12	US-10-072-012-491	Sequence 491, App
28	290.5	3.3	2802	9	US-09-808-602-81	Sequence 81, Appl
29	290.5	3.3	2802	10	US-09-800-198-69	Sequence 69, Appl
30	290.5	3.3	2802	12	US-10-072-012-489	Sequence 489, Appl
31	289.5	3.3	1364	12	US-10-282-122A-75309	Sequence 75309, A
32	285	3.2	2613	16	US-10-038-854-42	Sequence 42, Appl
33	285	3.2	2628	16	US-10-038-854-40	Sequence 40, Appl
34	285	3.2	2721	16	US-10-038-854-38	Sequence 38, Appl
35	285	3.2	2725	16	US-10-038-854-36	Sequence 36, Appl
36	284.5	3.2	2715	12	US-10-042-865-52	Sequence 52, Appl
37	284.5	3.2	2715	12	US-10-029-020-51	Sequence 51, Appl
38	280	3.2	1317	9	US-09-815-242-5118	Sequence 5118, Ap
39	280	3.2	1317	12	US-10-282-122A-43495	Sequence 43495, A
40	279.5	3.1	944	12	US-10-282-122A-47806	Sequence 47806, A
41	278	3.1	985	12	US-10-282-122A-43061	Sequence 43061, A
42	276.5	3.1	1438	12	US-10-282-122A-78360	Sequence 78360, A
43	273.5	3.1	1688	15	US-10-144-194A-113	Sequence 113, App
44	273.5	3.1	1737	9	US-09-808-602-83	Sequence 83, Appl
45	273.5	3.1	1737	10	US-09-800-198-71	Sequence 71, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-365-742-106  
; Sequence 106, Application US/10365742  
; Publication No. US20030204868A1  
; GENERAL INFORMATION:  
; APPLICANT: Collmer, Alan  
; APPLICANT: Alfano, James R.  
; APPLICANT: Cartinhour, Samuel W.  
; APPLICANT: Schneider, David J.  
; APPLICANT: Tang, Xiaoyan  
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING  
; TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF  
; FILE REFERENCE: 19603/4112  
; CURRENT APPLICATION NUMBER: US/10/365,742  
; CURRENT FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/356,408  
; PRIOR FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/380,185  
; PRIOR FILING DATE: 2002-05-10  
; NUMBER OF SEQ ID NOS: 209  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 106  
; LENGTH: 982  
; TYPE: PRT  
; ORGANISM: Pseudomonas syringae pv. tomato DC3000  
US-10-365-742-106

Query Match 10.5%; Score 932.5; DB 12; Length 982;  
Best Local Similarity 29.7%; Pred. No. 1.1e-65;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
QY 642 ANTLIDYELNQLQDNPFPVITTTDNGNOLNEFDGGRHVSOCLKSDGDKFYTI 701  
DB 31 SSTLGARYD-----DWNOR--CCVTTDDN-VQVTEYSDPIGSDVHK-----GPIQKT 74  
QY 702 HFQOQDEGRHHTSYSDYLTNGRQQTDPDKVHLSMKSVDNWCQIANTHWSYGVSEKIT 761  
DB 75 WKQSGDPEGR-----ISGRSET-----WLNL-----FKQPDRI 103

QY 762 VDPITATKQLOSNNVOTGKEVYTYTPSOQPIQITLFDAGHLQSCHTLTBDGWDV 821  
Db 104 -----TITAGTGRSHRWSRNLUT---TEQEL-----SRQFLYDGLGRC 144  
QY 822 RKEIDAIGQCTIYQYDYNRVQITLPDQTIYVNRKYAPFSDTLITDIRV---NGIS--- 875  
Db 145 TEQDALQOCTLPSYDNRWSVSTLADGSLVNRSYAFQSSSELATMLEVVHQNGTTRTV 204  
QY 876 LGQCTEDGLSRLOSODGGRWAYYSAGNDQCPSTVITPDQFIHYOYQPELDDAVLOV 935  
Db 205 AGTKQFDGERSVTQGTGRVSOQFNVDAGEMO-PRSRITAGLDNINFTYTRALTQDQIPSS 263  
QY 936 ASNEITQOFSYNPVTGALLKVAEGQSLTPYPSGRILKMNINDMKMSYLM-----T 989  
Db 264 TAPDETAKFDYDKTSARLIEATNPQGTTRYVDVHNQLTGETWDL--LQOAMETHQSS 321  
QY 990 LRGLNGYTDL-----TGFIQKISRDTGCRVTOIKDSIKTTLNVDLNRHIGSOVTDLA 1044  
Db 322 LLGRPIKRDLLKGEAGAEHETRYDITLGRIFNQSNERTITIDVVGQCKVATEDLQ 381  
QY 1045 TGHLMTTTFEFDGLNREIGRKLCDSSGHTLDIQOSWLKTOQLANRIVKLVGLQRTQYS 1104  
Db 382 AGTVIIMDEYDQGEILRTQASNOAALTQTWAVDGLLKTDRDQAGSPHLLHETFS 441  
QY 1105 YDSRNLNQYKCDGAECPDKYGHISIVTQNFYDIYGNITACHTTEADGTEHATKFEAN 1164  
Db 442 YDPRGLTLVWVGLSPLRDELQRETRQIFSDBLDNITICQTRFTDGTISERAAFKYS 501  
QY 1165 P-----TDPQOLTEVHTHPDMPDIRLKYDKAGRVINITDNGNENTFYDTLGLRL--- 1216  
Db 502 PGDKHKRCQLLSATYTPRKTPTDPTFSYDANGN--QLKDEHGN--SLHYDSOSRLLOV 557  
QY 1217 -ONGQG--SVYGVDPNLRVSKTKTLDCEL--VYRETNVNEVRNGEMIRLRTGTII 1271  
Db 558 AETGGAPIQYRYDGHQNLVATR-DONESEILRFEGHQLSSTVOEDORTQYLHLGEOPL 616  
QY 1272 AQQ--RASKVILTGDSQSVILTSQKQNSQAYSAYGKHSTANDASILGVNGERAD 1328  
Db 617 GQOIVDDAEQTLLLTANDQSVMGFEQOQLKAVYSAYGERHSEALLSTAGENGVE 676  
QY 1329 PVSQVTHLNGVRSYDPTLMREHPTSLSPFCAGINPYSCLDGPNRSDPSGH----- 1383  
Db 677 AANGWLLNGRFRAYNPLMRHSPDFLSPFRAEGVNPYTCNGNPLALRDPDTHDASQ 736  
QY 1384 -----LSWQ-----ANTGIGVGTIA-----GULLTIATGGMAIAAGGI----- 1416  
Db 737 TGLRPRDEGALPMQGGGDMGWGVGVGVVFTVLGVAATITATLTATPTVGPVTLGI 796  
QY 1417 -----AAIASTSTALAFG-----ALSVTSDITSIVSGALEDASPKASSILGHVS--- 1462  
Db 797 SMTASAAAASVSTGALLIVGTALTAATTANTVAIVNN-----DQTAGVGWHLGIAAV 851  
QY 1463 -----MOMGAAGLAESAIGKGTALA-----THLGPASDGENALLKSTSESR 1505  
Db 852 PVGLVGFAGAVVARAAAKVAAANAGTIGVRSVRIG-LAAAGARRTTISSAASAR 909

## RESULT 2

US-10-369-493-11243  
; Sequence 11243, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 11243  
; LENGTH: 1317  
; TYPE: PRN  
; ORGANISM: Methanosarcina mazei  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)-(1317)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-11243

Query Match 5.8%; Score 517.5; DB 15; Length 1317;  
Best Local Similarity 21.2%; Pred. No. 8.6e-32;  
Matches 334; Conservative 194; Mismatches 505; Indels 541; Gaps 72;

QY 39 TGLYNIQITLGHVGNGLGPTLPLTL--SYSPLNKTDIGFGIG--FNFGLSVYDEKNSL 94  
Db 56 TGSFYQYQDLSIPRG-----LPLTVSRYSNVDNRSLFGSGWTFNVMNLTVDNNGN 110  
QY 95 LSLSTGENYK--VIETDKTVLQOKKLDNLRPEKDKENCYRIHKSGDIEVLTFGNNN 151  
Db 111 VTVLGGDGHDTYILNPDGTVSRPLSVFDDL-----IKNSDETTLTKKDOT 157  
QY 152 AFDLKVPKLLN--PAGHAIYIDWNFEATOPRLNRIYDLDLGHDIPLNLEVOGLIKTI 208  
Db 158 KYNFSSEKLVNIYDKNGNQI---NFTYTCGQLTKV-TDASREL-ILAYDHG----- 206  
QY 209 LTLFPQGEYRTELRLFNRLQNLNHNFSNGNENPLTWSPGYTPIG-----KNGILGQ-- 261  
Db 207 -----HIISITDPMGRVMSYSYDDQGNLTCQKNPIGGKLS 241  
QY 262 -----WITSMTAPGG--LKETVNSNNQGHHPQSANLPVLPVVTLMKQVPGAGOP 311  
Db 242 VTYDENHMTSITDPRGNHPKNTY---DEKGVLSQS----- 276  
QY 312 AIQAEYSYTSNHYGGGNGIWNKLDNLYGLMTEYNYGSTRRYKDEGHDOIVRIER 371  
Db 277 -----NSLNATY--TFNYDS-ENRK----- 293  
QY 372 TVNNYHLLTSECKQNGYIOTTETAYVAILCHNFDQSOPOQLPKTKTETWRSADNSYS 431  
Db 294 -----TTETD-----PFCNKYTSFDEHFW-- 313  
QY 432 EITET-----TFDESQNLTKVIKDKTKOKIISPSTHWEYYPAGEVNCPEPYG 482  
Db 314 ELNETNQLGYTISAYDENGNRISVTNENSKTKLA-----YDANGNIK-TTNPLG 364  
QY 483 FTRFVKIIQTPYSEFPKDPKPEFIQVRYSLIGSQSHVTLKIEERHYSATQLINS----- 537  
Db 365 YSK-----SMTYDS-----KNNLISOTDELRHKTSFYDDNSNLKSIDALG 406  
QY 538 --TLPOINTKSELGRLLKQTECK-----GENCK-----TYSVVHKT 574  
Db 407 HETVFSYD-----KYQVIGETDNKKTKATFSYNNNGDQITITDANGKTSFTYDTVGRVT 462  
QY 575 -----YTKQDDTLQOQSHSITTHDNFTIHRSQVRSRYTGRL-----FSDTDTKDI 618  
Db 463 TKTDAGNRYTFQYDALDNLISITD-----PMGQTTSTNTMLLETIKVLQMLKVDXQSI 516  
QY 619 VTQMSYDKLGRLLTRTINSGT---PYA-----NLTLYDYELNNLODDNRPPFVIT 665  
Db 517 LTTLXISLXKXKQWQVEXLGNMMPLAIFWPPROTQKHKTSYDIDPLNQ-----VS 568  
QY 666 TTVNGNQLRNEFDGAGRHS-----QCLKSDSGDGKFTYTHIQOQYDE 708  
Db 569 VTNALGKTRNKYDAIGNKISITNAYGKSTRYSYNSLNLQVKTVMAMGK---VVRYNYDA 625  
QY 709 QGRHHTSYSD-----YLTNGRQOTDPPDKVHLMSKSYDNWNGQIANTHWSYVSEKIT 761  
Db 626 VGNLISITDENGHKINGYDSLRQVSVTDALRKTTRNKYDAVGNKISITNAYGKSTRYS 685

QY 762 VDPITLTATKQSQNSNNVQTCGEVYTTTPSQOPIQITLFDAGHLQSCHTLTRDGDWRV 821  
Db 686 Y-----NSLN-----QVKTVDAMGGVRYNYDAVGNLTIS-----715  
QY 822 RKTDAIGCTIYQDYNRYNIQITLPGDTIYNRYAPFSTDTLITDIRVANGISIGOOTF 881  
Db 716 --TTDANGRKTYGDSLNRQVSIINALGKTRNKYDAVGNKISSTDAWNR--LTKYSY 770  
QY 882 DGLSRLTOSQDG-GRVWAYTYS-A-NDQCPSTVITPD--GQFIHQY-----QPELDDAV 932  
Db 771 DSLNLVLKVTQDAMGGVRYTYDAVGN-----LISTDAKGKTYGYDVSLSLDRQVSIITDPL 825  
QY 933 LQVASENITQOFSNPVTGALLKAVAEGOSLTPYIPSGRLKMNINDMKMSYLWLRLG 992  
Db 826 GRTARENK-----YDAVGNKISSTDEGKTTISYGYDVLNRLTKYSYDDQKVSY-----873  
QY 993 LENGVTDLTGIOKISRTHGRVTOIKDSSIKTTLNVDLNRHI-----GSQV--TDLA 1044  
Db 874 -----NYDAVGNRLTMDKSHGTTAYKYDKLNRLLSVLNPDQCKVSYTYNK 918  
QY 1045 TCHML-----TTTVEFDGLNREIGRKLCDSSGHTLD---IQOSWLKTCOLANRIVKL 1093  
Db 919 VGNRVKMTYPPGKTTISYDAVNLIG--VIDSGHITSYAKNGNLKMTNTPNGV---973  
QY 1094 NGVLRQTEQYSDSRNRLNQYKDCGABCPDQKYGHSIVTONFTYDIYGNITACHTTPADG 1153  
Db 974 -----KTE-YSYDXANRL-----VELINKNTTQVSSYKYVTLDAAGNRLKVDEQLAEG 1020  
QY 1154 TED-----HAFKFPANPDPCQLTEVHTHPDMDPN--IRLKYPKAGRVINIT- 1199  
Db 1021 VESGDSSELKESQLTUTTYGY--DKYRLTKV-----DYPSENKTVSYKYDSMGNRISMTT 1072  
QY 1200 --DNHGNTENFTYTLGRLQNGQSVGYDPLNRLVSKQTDITLDCELY-YRETMVLNVEVR 1256  
Db 1073 NVDGIGSTISYKYDAADQLQSGNISYDKKGNLIKRVNSTQFMSYDE--ANRLK 1129  
QY 1257 N-----GEMIRLLRT--GETTIAOORASKVLLTGTDSQQSVILT 1293  
Db 1130 NVSEFVNTPPKSSYNFEYDGDGNRIKKTNGE-----NAQSTKYVLDINSALPQVLTE 1185  
QY 1294 SDKQNL-----SOEA-----YSAYGKHKSTANDASIL-----1320  
Db 1186 SDTKNTTCYVTGLISMTSENAYEYHYHGDGSGVRSLSDSGIIKNVTLYDAFGQVQK 1245  
QY 1321 -----GYNGERADPVGSVTHLNGYRSYDPTLMRFHTPPSLS--PFGAGGINPYSY 1369  
Db 1246 EIGTVDNNEFRFTGQMDDETGLIYLARY--YDPSVGRFITKDVIGRRVTTQSIINRYVY 1303  
QY 1370 CLGDPINRSDPSGH 1383  
Db 1304 TTNNPVLNVLDTGY 1317

RESULT 3  
US-10-156-761-7834  
; Sequence 7834, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7834  
; LENGTH: 2364  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7834  
  
Query Match 4.7%; Score 420.5; DB 14; Length 2364;  
Best Local Similarity 21.6%; Pred. No. 1.8e-23;  
Matches 320; Conservative 190; Mismatches 550; Indels 423; Gaps 73;  
  
QY 134 RIITHKSGDIEVLTFGNNAPDLKVPKLLNPAGHAIYIDWNFEATOPRLNRYVDDLDGDH 193  
Db 702 RKTGDDIDIVFVINFNRDLD-----NKVGS-----ELNF-----PRIKEIHGDL-----744  
QY 194 IPLLNLLEYOGLIKITLTLFPQKQEGYRTELRLNQLNSIHNFSLNENPLTWSFCYTP 253  
Db 745 -----GATTVSYGFAN-----ACDIDLHPAQA-----NTQDCWQ-KWTPE 781  
QY 254 GKNGILQMI-----TSMAPGGLKETVNYN-NNNOGHFFPQSANLPVLP- 297  
Db 782 GETDSKTGFKFLVTQVQVDFVTITNODGAPVMTTSYTYEDGAGWHF---TNDPLIKDE 838  
QY 298 -----YVTLMKQV-PGAGOPAIQAESYVTSNHYVGGSGGIWNNKLDNLYGLMTY 347  
Db 839 DESWTDWRCYQEVQVTTGAGACQK-----TKSWLYRGLSGDRTSKAD-----881  
QY 348 NYGTESTERYKQKEHQIVRIERTYNNY-----HLLTSECKQONGYIOTTETAYYA 399  
Db 882 -ASATKTVTVDDGQ-----NNYDSDLSGRILSTSLRDDTG--TSHERTYHK 927  
QY 400 IIGHN---PDSOPSOLFPLKXTETWRSADNSYRSEITETTTDES-----GNPLTKVI 449  
Db 928 YNDHNTAQDGLDARFVREKETTTNTKVSQWREHTVETETEDDTGEGASTTFGLPMR---984  
QY 450 KDKTKQKIISPSTHMEYPPAGEVDN-CPPEPYGTRFVKIIQTPYSEFKDDPEKFTQ 508  
Db 985 TDDWGQSSVS-----DNRC-----TTYGAYNTD-----1008  
QY 509 YEYSLIGSOSHVTLIKIERHYS-----ATOLLNSTLFOYN-----543  
Db 1009 -NYDSTGAQRWTVLQDQVKHYSVGCSSIAQSKQGYASTLYDNATSVANKPVDGNPTES 1067  
QY 544 -----TDKSELGRLLKQTE-----CPKGENGKTSYVVRHFTVTKQDDTLQ 584  
Db 1068 RYTKSGSVRSWTCYDDAGRWMSDGHNRILTLYSPANTWPLNGVIVITPPDGPALP 1127  
QY 585 SHSIITHNFTIHRQVRSRYTGRFLSDTDKDIQVQMSYDKLGRLL-----631  
Db 1128 AHTALTSTAWT-----SREWGKFTSIQDANGNVTKMSLDAAGRLVEVRPTETGSSPS 1180  
QY 632 -----TRTLNSGTPYANTLTLYDYELNNLQDNRPPFVI 664  
Db 1181 MKFSTIPTSTNSAGVDAVDGYPHVAITHVLOSQTTLSSHAYVDGLGRAR-ETQTPMGN 1239  
QY 665 TTTDVNGQLRNEFDGAGRHYSQLKQSDGDKFYTIHQOYDQGR-----HH 713  
Db 1240 GYDAATGNEVEN-----RQVSVTRYDSAGN---VTGTSVFRNQGTAGSGPSSAKVED 1290  
QY 714 TSTYSIDLYTN--GRQOTDPDKVHLMSKSYDNWQCIANTHWSYGVSEKIVTDPITLTATK 771  
Db 1291 LPSYTDLVLDWAGRAITS-----RLQVNGASQDAGRVTPTTYDGDFTSVKKNVDAADTYD 1346  
QY 772 QLQSNNNVQTKGEVT-----TYTPSQOPIQITLFDAGHLQSCHTLTRDGDWRVRET 826  
Db 1347 VYGVQSVKVEHTGSATVYTYTAKDELKIT--DPRGN-DTSYTYDWAQORTATDSD 1403  
QY 827 AIGQCTIYQDYNRYNIQITLPGD-TIVNRKYPFSTDTLITDIRVANGISIGOOTFGLS 885  
Db 1404 AGVST--EYDKNGQIKSITNSGKTVLDYGDALGRKTA---VRSGDTLEAATWBDGLN 1458  
QY 886 RLQSQDQGRVWAYTYTSAGNDQCPSTVITPDQGIHQYQOP-----ELDDAVLQVASENI 940

Db 1459 -IT---GGK-GQITSTGRDNGNTYTTKTGEF-DVRGRPLKTTVTIPDVKGLAGS-Y 1510  
QY 941 TQOFSNPVTGALLKAVAGQSLTPIYF-SGRKXENINDMKMSYLWTLRGLNGY-- 997  
Db 1511 TTGFTYD-----AADHVSVAIPAAGGLPAETVN-----AYDGYR 1547  
QY 998 -----TDLGTTQKISRDRHGRVTOIK-----DSSIKTLLNYDDLX--RHIGSQV 1040  
Db 1548 PLRLQALQTYIRSTGYDAYDLTDRSYGDSVLPFGIGAAAQRTYSYDSDNGTRELKSA 1607  
QY 1041 TDLATGMLTTFVEFDGLNREIKRLCDSSGHTLDIOQSWLKTQOLANRIVKL-----NGV 1096  
Db 1608 T-----TTT-----LNKVSERQKDT-----YTYDLAGKLTLEQASGQ 1642  
QY 1097 LQTEQSYSDNRNL-NQYK-----CDGAECPDKYGHISIVTQNFYDIYGNITACHT 1148  
Db 1643 TAQSQCFLYDDQARLTNAVYTHTTTGICADKTKTASDFKGTAPYQYATYDRLGNLQITN 1702  
QY 1149 TFDAGT-----EDHATFKCANPTDPCQLTEV-HHHPDMPDNIRLKYDKAGRVI 1196  
Db 1703 TDSAGTATLHXYLPGYDDTGTTTANANQPHGVKIDHKGTGTTKSDQLYYFDDGTMK 1762  
QY 1197 NITDNGHTENFTYDLGRLO-----NGQGSV---YGYDP-----LNLRLVSKT-----DTL 1240  
Db 1763 QRVE-FGITTDTWTTPQGLEAVKTKSGSGSELTRYAYDADGNILVTRTPQETVASIDGM 1821  
QY 1241 DCELYYRETMNLNEV-RNGEMIRLLRTGTTIAQORASKVLLTGTDSQSVILTSKDN- 1298  
Db 1822 ELATNGTIVTATRVASGTATVARTTEGITATN--GKVYLMADQASTOLAVDASTG 1879  
QY 1299 -LSQEAYSAYGRKHS--TANDASILYNGERADPVSGVTHLNGVRSYDPTLMRHTPDLS 1356  
Db 1880 ASSTRRYTPFGERSGTLPTGDNGLGKTEDTSTGLSL--GARAYDNLNGLFSLPDPL 1937  
QY 1357 -SPFGAGGINPYSYCLGDPINRSDPSGHLWQAWTGIGMGIAG 1398  
Db 1938 ATPYAPQNLNAVSYATNPNISYSDPSG-LCRREDICGDGYPVAG 1979

## RESULT 4

US-10-156-761-7990  
; Sequence 7990, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIOYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156, 761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7990  
; LENGTH: 820  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7990  
Query Match 4.4%; Score 395; DB 14; Length 820;  
Best Local Similarity 23.8%; Pred. No. 3.2e-22;  
Matches 242; Conservative 120; Mismatches 375; Indels 280; Gaps 50;  
QY 501 DPEKFIQVYSLIGSQSHVTLKIEERHYSATOLLNSTLFCYNTDKSELGRLLKQTEC-T 559  
Db 26 DTEGKTTVFRYDANGRVOKIT-----TASGRVTVFTYD-DANRVTSMLRGTGFSN 74

QY 560 KGENGKTYVWHKFTYTKQDDTLQOQSHSIITHDNFTIHRQVRSRYTGRLFSDTDTKDIV 619  
Db 75 DGHGTPTW-----TYAYTSDSVTAAGTTATDPTET-HATKYQHDSDGVQSDVTDAMGHK 127  
QY 620 TOMSDYKLGRLLRTR--LNSGTPYANVLTVDYEL-NMLQDNRPPFVITITDVGNOLEN 676  
Db 128 RSTKEDANHSIDTSDAMSGSTTFCNVDTYGFNTRNLE-----TITQPTGGKTVN 178  
QY 677 EPDQ-AGRHVSQCLKSDGDGKFYTIHTQQYDEQGRHHTSYSDYLTNGRQQTDPQKVLH 735  
Db 179 HQTIAGADVPKDSINPDGKTDFT-----YDAVGN-----TW 211  
QY 736 SMSKSYDNWGOIANTHSGYGVSEKIITVDPTLTA---TKLOSNNNVONGKEVITYTPS 752  
Db 212 SVAQTGTGGNVSYTY-----NPASPTCGGFAGORCTEKTMTAAKTVTT--- 256  
QY 793 QOPIQITLFDAGHLQSCH-----TLTRDGDWRVRKETDAIGCTIYQYDNYN--RV 842  
Db 257 -----TFHYDSAGNLDVTVPAPLAKITYTYDALGRKTKVTDARGVTVTYTYDNRDTRI 311  
QY 843 IQITLP-----DGTIVNRKYAPESTDTLITDIRVNGISILGQOQTFDGLSRLT--OSQ 891  
Db 312 VDTNKAARVEWYDGDGNLTOR--TDSGTIKYD-----FDFLQRETIIRTLQ 356  
QY 892 DQGR-VMAITYSAGNDQCPSTVITPDQGFHYQYQPELDDAVLQVASNEITQOQFSYNPVT 950  
Db 357 DGSQTLAY-----TPSGNVVYQ-----DPAG-----TVDYTWNEV- 398  
QY 951 GALLKAVAEQOSLPIYPPSGRLK--MENINDMKMSYMLTLRGLNGYTLDTGTTOKIS 1008  
Db 399 -----NKLAEKDPAGRVITYKYNNDVKTIT--TYFG-----GTQCKVD 426  
QY 1009 RDTGRVTOIKDSSIKIT--LNYDDLNRHIGSQV---TDLATGHMLITTFEFDGL 1058  
Db 427 PUNSSRPKTIKTTSPKGTVDLAYSQYGTSGTTEGSKIRSDNVIG--LKTITYYDGA 484  
QY 1059 NREIGRKLCDSSGHTLDIOQSWLKTQOLANRIVKLNGVLORTQOYSDSNRLNQYKCDG 1118  
Db 485 GRFSYAE--EKKGTLLN--SSWLYCYDLAGNL-----TSQGAAGCPGATTTVND 531  
QY 1119 AECPTDKYGHISIVTQNFYDIYGNITACHTTFADGTEDHATFKFANPTDPCQTEVHHHT 1178  
Db 532 AQLTAKNGS--TTNWSYDKIGNETAGAST-PEGTRAEKW-----TDHSQLTSL--- 578  
QY 1179 PDMPTNRLKYDKAGRVINITDNGHTENFTYTLG--RLONGQGSYGVDPDLNRLVSQK 1236  
Db 579 -----TVGKK--TYTQYGSTDQSERIRLGDYTFHNGPIGLAG----- 614  
QY 1237 TDTLDCELYYRETMNLVNEVRNGEMIRLLRTGTTIAQORASKVLLTGTDSQSVILTSK 1296  
Db 615 TSTAGVDTGFNR-----EPGQTLNSMTTGKAYCYLTDALGSLVIALTDE 658  
QY 1297 --QNLQEAAYSAYGKHKSANDASILGN--GERADPVSGVTHLNGVRSYDPTLMRHT 1352  
Db 659 TGAQNTVYVSPRGVTRASTSEKNPQPRFAGYQD-VTLGYHYGARY--YDNIORFNS 715  
QY 1353 PDSLSPFGAGGINPYSYCLGDPINRSDPSGHLWQAWTGIGMIGLITITATGGMALAA 1412  
Db 716 PD---PSQEE-NPFLYAEGDPVNRIDPNGLFSDALDTGSDIFGVWTVGCLAGVTAAG 771  
QY 1413 AGGTAAATASTTALAFGALSALTSITSDITSVSCALEDASPKASSILGWSMGMGAAG 1469  
Db 772 TGSIAAAA-----VGVVVGAGVSAGVGLAVVVGSCALG-GAAG 810

## RESULT 5

US-10-156-761-7572  
; Sequence 7572, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO

```

; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7572
; LENGTH: 1250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7572

Query Match      4.4%; Score 392; DB 14; Length 1250;
Best Local Similarity 21.1%; Pred. No. 1.2e-21;
Matches 271; Conservative 187; Mismatches 452; Indels 368; Gaps 66;

QY 254 GKNGILGOWITSMTPAGGLKETVYNNQGHHPQ-----SANLPVLPVYVTLMKQVPG 307
DB 59 GHEGVAEGGLAK-TARNAVEDAEDASQKAKICEKDPVDVATGRVLPQI-----DVSL 113
QY 308 AQOIPAQAEYSYTSYNNYGGGSGNGIWNKLNLY-----GLMTEYNYGSTES 354
DB 114 PGQLPLVVKRGFESSYRLGGWFGPTWSSTLDQRLVDTAGVVVGEDGLVLAIPH-----PA 170
QY 355 RRYKDKEGHDIQVIERIYNNYHLLTSECKQNGYIOTETAPYVAAIIGHNFDSPQSOLF 414
DB 171 PGVPTLPSHGPRWLDRTDGGYTL-----TDQKSGHIR----- 203
QY 415 PKTETETWRSADNSYRSSEITETTFDESCNPLTKVKKTKQKIISPSTHNEYVPPAGEVD 474
DB 204 -----HFVDRSDTLAVLEQLDDRNWNIT-----FEY-----DAD 233
QY 475 NCP-----PEPYGTFREX-----KIIQTPYSEFKDDPEK-FIQVRYSLIGSQSHVTLKIE 526
DB 234 GAPRYLTHSGGYRLRISTEAGRVTAHLASAAVDGGQDLIRGY-----TDGHL----- 283
QY 527 RHYSATOLLNST--LFOYNTDKSELGLLKQTE-----CTKGENKGYTSV 569
DB 284 -----TEVNSSGRPLQVCD--ELGRITSWTDNSHFSYAYDDEDRCTH-QSGAAGHL 335
QY 570 VKFTYTKQD-----DTLQQSHSIITHDNFTIHRSQVRSRYTGRLFSDTDKDIVTOM 622
DB 336 RSTFAYGAVDPGTGAHTTTVDSYGQTH--YLLNR-----RCQVIAETDALGAVTRY 386
QY 523 SYDKLGRLLTRTNSGTPYANTLYDY-ELNNLODDNRPPFVITTTDVNGNQLRNBFDGA 681
DB 387 QDRYNELLSQI-----DPLGHTTSFRYDDAGNLVAATRP-----DGREARAEYNAL 433
QY 682 GRHVSQCLKDSGDGKPYTHIOYDQGRHHSSTYSYLITNGRQOTDPPDKVHLSMSKY 741
DB 434 GLPVK--LVNPDG-----TITROTDFRG-----NUTSVTDES--GQTRFGY 472
QY 742 DNWQIANTHWSGVSKITVD-----PILTL-----ATKQLQSNNSNVQCKEYVITYTPSQ 793
DB 473 DEGRLLTSMTDPLGHTTGVCDRAGLPLTVTDPLGAVTRYERDAFGRTTA--ITDPTGAT 530
QY 794 QPIQITLFDAGHLQ-----SCHTLTRDGRVREKETDAICQCTIYQYDYNVRVIQT 846
DB 531 TREWTV-----EGHLSRTADPGSESMTYDGENCTSHTDPPGVGSLFEYTHFDLPTART 587
QY 847 LPDGTIVNRKYPFSTDTLITDIRVANGISLGQOTFDGLSRLTOSQDQGRVWATYTSAGND 906
DB 588 GPDGV-----RY-EFEHD--TELR-----LSQVTNPH--GLTNVAYDAERG 624
QY 907 QCSTWITPDGQIHHVQYQPELDDAVLOVAS--NEITQQFSY-NPVTGALLKAVAEQQSL 963

625 LVAETDF--DNRILTVEY-----DFAGRLASRTNAGLQMIAFERNELGQIVRKDAAGQAT 677
964 TPIYYPGSRGLKVENINDMKMSVLTLRGLENGVTDLTGTIQIKISRDTHGRVTQIKDSSI 1023
678 TYAYDFDQLAQ-----TGPDGTALTILDRHGRMISEAVNGR 716
1024 KTLINVDLNRHI-----GSQVT-----DLA-----TGHMLTTTTFDGLNRIRGRKLCD 1068
717 ALTYDYDELGRTRTRTPSSATTIWSYDVAGRETSMTGSGRSIDFVYDEAGRELGRFF-- 774
1069 SSGHTLDIQOSWLKTQOLANRIVKLANGVLQRTQEQSYSDSENRLNQKCDGAECPDKYGH 1128
775 --GH-TLHSPDALGRLTSQSV-LGPAGRUHQRY-----TTRAD-----GH 814
1129 SIVTON-----FTYDIYGNIPACHT-----TF-----ADGTEDHATFKFANP----- 1165
815 LIGIEDQLSGKRFRDLDLAGRVTAHAVNWTETIYDAYVGNQTSAGSWPAGHPQOEALGNR 874
1166 ---TDPCLTEVHHTHPDM-----PDNIRLYKDYKAGRVINIDNHNGTEN 1207
875 TYTGTSLTRAGQVRYEHDGLRIILKQTRLGRKPTWRYTWDADRLTSVTPDGTWR 934
1208 FTYDTLGR-----LQNGQGSYGYDPLNRLVSKQKTDLDGCELYYRETMVNEVR---N 1257
935 YTYDPLGRRTAKRLAEADGETLV-----ERVDFTWDGTVLCEQTTSPDLPHQVTLTW 988
1258 GEMIRLLRTGERTIAOORASKVL-----LGTDSQQSVILTSQKONLSQEAAYSAY 1307
989 HQGLRVPVOTERIVAADAPQOEIDRSFFAIVTDLVGTPSE---LLDERGEIVMRTRATLW 1045
1308 GKHKSTANDASI--LYNGERADPVSGVTHLNGVRSYDPTLMRPHTPDLSLSPFGAG-GI 1364
1046 GSTTWAKNSTAYTPLRFPQYYPETGLHY--NYFRHYDPEIARYVTPD---PLGLGPAP 1100
1365 NPYSYCLGDPINRSDPSG 1382
1101 NPAAV-VPNPHWADPLG 1117

RESULT 6
US-10-156-761-7751
; Sequence 7751, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7751
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7751

Query Match      4.4%; Score 391.5; DB 14; Length 2386;
Best Local Similarity 20.1%; Pred. No. 4.1e-21;
Matches 379; Conservative 214; Mismatches 688; Indels 601; Gaps 93;

QY 163 NP---AGHAIYIDWNFEATQPLNRIYDDLDGHDIPILNLEYOGLIKITLTLFPQKEGY 219
```

Db 698 NPDGTTCKTLMWDYIQRKT-----YGDGDDVLVPVINFRTDL-----DNKVG- 740  
Qy 220 RTBLRFLNQLSINHFSIGNENPULTWSPG-----YTPIQNG 257  
Db 741 SABLNF--PRIQBIHG-DIGATTKYSYGFANACDIDLHPTQASNTQDCYWKWTFEGETE 797  
Qy 258 ILQOMITS-MTAPGGIKETVYNNQGHHPQSANLPLVPVILMKQVPG-----AGQ 310  
Db 798 SKTGFKKFLVKVEVDPTV--TTWQDG-----APVMTSYIEGAGMRITGD 844  
Qy 311 PAIOAE-----YSYTSNHYVGGSGNHNKLDNLYGL-----MTEYNYGSTESRYKD 359  
Db 845 PLVKDESDSWDRGQYQETQVTTGADTGQTKKYLWYRGLDGDRTSKTDTSATKTVTND 904  
Qy 360 KEGHDQIVRIERTYNN-----HLLTSECKQONGVIOQTETAYVALIGHN---FDSQ 408  
Db 905 GDG-----NNYADHANLAGHTLSTSLRDDTDI--SHERTYHYWNSHNTAQYDGL 951  
Qy 409 PSQFQPKYKTEWRADNSYSEITETTFDES-----GNPLTKVKKKTKOILSPS 461  
Db 952 PDAHFVRESKSTNTKISGMRHVEHVENYDDSEASTTFLGPMR-----TDDWGOSNYS-- 1006  
Qy 462 THWEYPPAGEVDN--CPPEPYGTFRVKKIQTTPYDSEFKDPEKFIQYRYSLLIGSQSHV 520  
Db 1007 -----DNRC-----TTGRAYNTD-----NYDSTGAQRWT 1031  
Qy 521 TLKIBERHVSATQLNSTLFPQNTD--KSELGRLLKQTECKTGNGKTYSVVHKFTYTKQ 578  
Db 1032 VQDQVKKHVSUG-----CSSIADSNQDGYTSTLYDNATSIDANKPVDGNATEV---RITYTK- 1084  
Qy 579 DDTLQSHSITTHDNFTHRSQVRGRYTGRLFSDDTDKDIQVMSYDKLGRLLTRTLNSG 638  
Db 1085 -----AGYRSRWG-----YDKAGHIVWS--EDG 1107  
Qy 639 TYPANLTLYD-----YELNMLQDN-----RPPFVI-----TTTDVNGNQ 673  
Db 1108 XHNRSTIYSPANTWPMNGITSPDPDGTATARGPLTFTEWTSFPAWTFITTSKDANGNI 1167  
Qy 674 LRNEFDGAGRHSQCLKSDGSG-----KF-YTHITQO-----YDEGRHHTSTYS 718  
Db 1168 TKVTLDAAGPEVEMKPIETGSSPSMKFSYIPISTNSAGVPSDAGPYVASHTLQSGS 1227  
Qy 719 DYLTN-----GR-----QOTDPKVLHMSKS-----YDNMGQIA----- 748  
Db 1228 TYLVSHAYTDGLGRARETQTPLPSPVDVDPATQIVPFROVAVTRYDSAGOVGTGASAVFRNQ 1287  
Qy 749 NTHWSYGVG-----EKTVDPIITLTATKQLQSNNSNVQCKEVITY-----TP 791  
Db 1288 GTAGSGGPPSPQSDLPYSYDVLWDAGRTVSSQIQVKGTPQKAGRYDTSYLGDTYSVIP 1347  
Qy 792 SQOPIQITLFDAG-----HLOCHTLTRDGW-----DRVKETDAIQOCTIYQYDNYR 841  
Db 1348 VDSATDITYDYGVQSVKVEHTASAYTTAYGTAKGELAQLITDPGNNLTLYDWAAQ 1407  
Qy 842 VIQITLPDGTIVNRKY---APESTDTLITDIRVNGISLGQOFTDGLSRLTQSQDGG---R 895  
Db 1408 RKTDDPDAGLSSEYNGQVSQITATNDVQVTLTVG---YDNLBRATSVRSAGADELA 1464  
Qy 896 VNAY-----TYSAGNDQCPSTVITPDQGFHYQYQPELDDAVQVANSNETQQ 943  
Db 1465 AMWDDPAAATGKGKITSAVSDASGNTYTTKTGK-----DERGRPLNTTVTL 1514  
Qy 944 FSNVPTGALLKAVA--EQQSUTPIYP--SGRLKMNENI-----NDMKQVSYLWTLRGLENG 996  
Db 1515 TTVNGLAGDYTSVTVYDAADHITSVSPYPAACKLAACKVTTYTDYDQPTRLTSLG--- 1570  
Qy 997 YDITLCTIQKISDTHGRVQIKDSIKITLN-----YDNLNHHIGSQ-VTDLATG 1046  
Db 1571 ---GTAYDNTTYDAYGLRVE-RDYGAEEFGNGQAORQYGYDSN---GTRMLRSIAT- 1622  
Qy 1047 HMLTTTVEPDGNRIBGRKLCDSGGHTLTDIQCSWLKTKQOLANRIVKLVGLQRTQYQSYD 1106  
Db 1623 ---TTTINDLVSEAQKDTYLVNTCKLTRE-----QASGQTAQSQCLRYD 1666

Qy 1107 SHNRLN-QYK-----CDGAECPTDKYGHISIVTQNETYDIYGNITACHTTFADGT--- 1154  
Db 1667 DOSRLTLAYTHHTAGNCADTTKTSDFKTSFYQTGYTDRLGNLQSVTDTNAGAAATR 1726  
Qy 1155 -----BDHATFKFANPDPCQLTEVHHHTPMDPNIRLKYDK-----AGRVINIT 1199  
Db 1727 DYLYPGYDAGTWTAN-----ADQPHGVR-KINKVSAGTTTAACTTYTVA 1771  
Qy 1200 D-----NHGNTENFYDYLGL-----QNGQSVYGYDPLNPLYSQKTDLTDLCEL 1244  
Db 1772 DGAMKQVREGSTTDTYWSRLGRLATVKTTSGLSDLTRTYDASGNLLVRITP----- 1825  
Qy 1245 YVRETMV---NEVRNGEMIRLLRT-----GETIIAQORA-----SKVLLTGTDSQOS 1289  
Db 1826 --QETVASIGGTETLRTTDTGIGSATATRYVSPGATTVAMRTDGNNTVNGKITVLMGDTQAS 1883  
Qy 1290 VILTSQKQ--NLSQEAYSAYGKHS--TANDASILYNGERADPVSGVTHLNGCYRSDPT 1346  
Db 1884 TQIAVDAATGATATRRYTPFGDERSGSLPTGTNHHGFLGKTEDNTGLSL--GARYDPS 1941  
Qy 1347 LMRHTPDSLS--PFCAGGINPYCYCLGDPINRSDPSCGLSWQA-----WTGICMGIA 1397  
Db 1942 LGRFLSPDPLSTPYDPQNLAYSYSNGNDPINYSDPSGLIKLNSDGTQCSGDMQCGPGTV 2001  
Qy 1398 G-----LL-----LTATGMAIAAAG-GIAA 1418  
Db 2002 GGGGDTPTTAPTVQDQLVDLLPREENGWDADRLAQVWHYQTTQGGYWDAPVGGDRT 2061  
Qy 1419 ATASTTTL--AFGALSVTSDITSIV-----SCALEDASPKASSILG 1459  
Db 2062 GWACFGRTACSEAFVWKETHDFAKAKRVAATFCVENPKCGADNGAYDSMKSEASAVPI 2121  
Qy 1460 VWSMGMAAGLAEBAIKGTTKLATHLCAFAEDGENALLK--STSESRILKWG-----V 1510  
Db 2122 LLAGEMGAA-----FSKVLKARGCSFKPTTRVLKMDGKTKPLGKIPGDLVEAADP 2172  
Qy 1511 TRSLDREI-----VRNEEGQVVKDHS-----RGYTDNFMKGEOAILV 1548  
Db 2173 TSGHREVRETVAVLHNDLDDVLSIRGLDRIQTLHTTARHRIWDDTAQVMEQAGRLI 2232  
Qy 1549 HGDYDGLYHTEGKH-----NGKGYTRTEPEQLVDYLDKNNIVDLTQGGDKP-- 1597  
Db 2233 TGHK-----VNTSGNQHATITSVLAQRAADMYDLTVEGLHTY-----YVLAGEIPVL 2280  
Qy 1598 VHLSCYKSS-----GAADKMAKVINRPVIAYSNKPITISQGLARIERKDFPKSTYHS 1651  
Db 2281 VENGSCWSSTNRKTSVKNAFGHKKH-----KSBFPNLNNAKEYVEAGTDFLRST--- 2330  
Qy 1652 YDPKILIGRTEKTVKPKTRP 1673  
Db 2331 -DPS--VLTRFRANGDVIRNP 2349

## RESULT 7

US-10-282-122A-46565  
; Sequence 46565, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Chlesen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A

;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR APPLICATION NUMBER: 60/191,078  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR APPLICATION NUMBER: 60/206,848  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/207,727  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 46565  
;; LENGTH: 2234  
;; TYPE: PRT  
;; ORGANISM: Bacillus anthracis  
US-10-282-122A-46565

Query Match 4.3%; Score 386; DB 12; Length 2234;  
Best Local Similarity 18.7%; Pred. No. 1e-20;  
Matches 339; Conservative 236; Mismatches 633; Indels 608; Gaps 76;  
  
QY 21 PTOANNFSAVSGVDPR-----TGLNYIQTILGHV-----NGNL-----57  
DB 748 YSKVGHAESEISHFVPRFSODSGFICMFDYNASIPVLNGKVNATNGFIMSEKDITLSG 807  
QY 58 -GPTLPPLTSLSPNKNDIGIGIFNFGLSV---YDRKNSLLSL-----TG-EN-102  
DB 808 RGPDSVVERTNSQSKVGLFGTGWSSGLSEERWADGNGNLLIISTDGANITRTGDNK 867  
QY 103 -----YKIETDKTVKLOQKLDNIRFEKDLKENCYRIIHKSGDIEVL 145  
DB 868 YQAPTGYLEIKQVSGGYEIKDQTV-----TFYKSGDAQQR 905  
QY 146 TGFNNNAFDLKP-----KKLLNPAGHAIYIDWNFEATQPLNRIYDDLDGHDPL 196  
DB 906 IEYTKDKYGNITTYEYDGAERLSKVNASGKELVL---QYDGNKKAAVLI-----GPDNKT 959  
QY 197 LNFYQGLIKTILTFPGQ---KEGYRTFLRNLRLNSIHNSLGNENPLTWSFGY-----250  
DB 960 ITFNYDGLLVSSITPFGKVKYGYD-----NGVLTSIYDPQHTDAKPKYKTSYAYENDR 1013  
QY 251 -----TPIKNGNL---GQWISMTAPGLKETVYNNQGHHPQSANLVPILPVTL 301  
DB 1014 LVKVTPLKATLAVNNGSKVELTLPNPKRKTYTYND-----AGNFPV-----1057  
QY 302 MKQVPGAGQPAIOAEYSYTHNYVGG-----GSGNIWNKLDNLYGLMPEYN- 348  
DB 1058 -KTVEDVGRNLNLTTSYEYNNANLVKITPKNQETATYDGNVNTSVDMGTEKEFYNK 1116  
QY 349 -YGSTESERRYKDE-----GHDQIVRIERTVNNVHLTSECKQNGYIQTITETAYAI 400  
DB 1117 DNGIILKATDNEDRKTTVAYGANTSVSDQGAN-----TSSVIHHDQYGNPIETSKELS 1171  
QY 401 IGHFDSQSOFOPLPKTKTETWSADNSYSEITETTFBESGNPLKVKIKDKTKQKIIISP 460  
DB 1172 AGGNLQNPFS-FEM--NGIEKWVKVDNNNGSIS-----KDATPACGLGGESSLKITTK 1223  
QY 461 STH--WEYPPAGEVENCPEPPEYGFTRFVKKII-----QTPYDSEFKDDP 503

DB 1224 ATNNDWGYIAAIOEVTLPEPNTTYTLQGMVKTDLVNGAAFFNVQSLNENGAGIDGGWHDTR 1283  
QY 504 EKFIQYRYSLSQSHVTLKIEERHYSATOLLNSTLFOYNTDKSELGR-LKQTECTKGE 562  
DB 1284 HNKVQGTSDWNRQ---VTFKTEQ-----TRKVIYLOVENGGSATSGAWDFDKIUEKGE 1337  
QY 563 NGKTSYVVKHFTYTKQ--DDTLOQ-SHSITTH-----DNFTIHRQSVRSRYTGRLFS 611  
DB 1338 VSSFPNVLNSSFENWPGFVQWVRSCSQHCERNVDVSDSFTGHSSIVMER-----S 1391  
QY 612 DTDTKDVIQMSYDKLGRLLTRTLNSGTAYANTLTLYDYELNLODNRRPFVITTTDVG 671  
DB 1392 EYCPNDI-----GYNRVILNOKKAEVTVLTAMSKSENVND-----1428  
QY 672 NQLNEFDGAGRHVSQCLXSDGDKGFYTIHTQYDEQGRHHTTSTSYDLTNGRQOTDP- 730  
DB 1429 -----APDKLSKDYAVLAETYYQDG-----TVVNYTSPFSTNDW 1464  
QY 731 -----DKVHLSMSKSYDNWGOIANTHWSYGVSEKITYDPIILTKATKQLOSNS 777  
DB 1465 NRSAAVIPAKKPIQKIEIFLLFRKNKG-----KVFDDIILLEGNALIKNE 1511  
QY 778 NNVTGKEVTTYTPSQPIQITLFDRAHLQSCHTLTRDGDWRVRKETDAIGQCTIYQYD 837  
DB 1512 YD-NDGNVVATY-----DEEGQK---NFTYDASGNKKSETDEKNTKLYDYN 1555  
QY 838 NYNRVIQITLPDGTIVNRKYAFSTDTLITDIRVNGISLGQOTFDGLSRLTOSQDGRVW 897  
DB 1556 KNLATKVTLKNGTSVNYRYD-----VTS-AT-----DSRGGKID 1702  
QY 898 AVTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQOFSYNPVTGALLKAV 957  
DB 1577 -----HNGNTEKSVMPFGKGTQTKHEYVDVKNVYIDALNRRIEN--TYDENANKIKTKM 1631  
QY 958 AEGQSLTPIYPSGRLKVENINDMKMSYLTWLRGLENGYTDLTGTIQXISRDTHGRVTO 1017  
DB 1632 PNGSILESVDYADRVRGK-----RNGKDSFT-----FERDQNGQVTK 1670  
QY 1018 IKD--SSIKTLNLYDNLNRHIGSOVTDLATGHMLTTFVEFDGLNREIGRKLCDSSGHTLD 1075  
DB 1671 VKDLVNGVERTYDKADR-----VTS-AT-----DSRGGKID 1702  
QY 1076 IQCSWL-----KTQOLANRIVKLVQLRTEQSYDSRNRLNQ--YKCDGAECPDCKY 1126  
DB 1703 ---WAYHDKANSKTEKLEQVTVQGY---TNKVSYD-YNTLDQNIIVTDSQ-----1748  
QY 1127 GHSIVTQNTYDIYGNITACHTTFPADGTEDHATFKF--ANP-----TDPQLTEVHHT 1177  
DB 1749 -----TYRFDYDDQGNV---RTYTAGNGSGSTFNVDQANKIKDLVVGTSNLSILSERYE 1799  
QY 1178 HPDMFPDNLRLKYDKAG-----RVINITDNHGNTEFNFTYDTLGR-----LQ 1217  
DB 1800 YDQSGNRTKIKHEGAGKVTETNFVYDPIINQLLNEVLNPGTTSYTYDGFGRNRSVKVIE 1859  
QY 1218 NGQGS---VYGVDPNLRVSKTDPLDCEL-----1244  
DB 1860 NGKETKTAATFNEGNQLVKFGNESLTYDVANGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919  
QY 1245 -----YVRETMLVNEVNGEMIRLLRTGETI-----IAQQ 1274  
DB 1920 APATKYDEDNRRTEKXVNGQVTRYFYDGSINPLIYETDNGTVLRQVYVYVADGARLAWK 1979  
QY 1275 RASKVLLTGTDSQOSVI--LTSKONLSQEAAYSAGK-----HKSTANDASILYNGERA 1327  
DB 1980 AQGQTLVYHYNPRGVDVAVANTQDKEVATYEDAWGNVLTSDTKGIAAD-NPFYVAGVY 2038  
QY 1328 DPVSVTHLNGYRSYDPTLMRPHTPD-----SLSPFCAG-----GINPISYCLGDPINR 1377  
DB 2039 DKEIGMYL-----IARYNPEHGVFLSVDDPDGDEDDPVTWNGYTYADNPNVMM 2088  
QY 1378 SDPSGHLSSQAWTIGMGIAGILLATATGGMAIAAGGIAAAIAASTSTALAFAGLSVTS 1437  
DB 2089 TDPDGKAW-----LVPVVIAGAWAARFGAKYAI-----RYGAKYKKA VKS 2131

QY 1438 DITSIVSGALEDASPRASSILGWVSMGMAAGLABSAIKGGTKLATHLCAFAEDGENALL 1497  
 Db 2132 -----GWY-----GKVKASGWNKGSIAQKIPRIHKVGR---I 2163  
 QY 1498 KSTSESRIKMGVTRSLDRIVRNEEGQVVKHSGRYTDFNMKGGEQAILVHGDKDGFPLY 1557  
 Db 2164 KGDNDKGGYGVYIYTK-----TKGKRTYS-----SFEFHTPHNGHY 2203  
 QY 1558 HTEGNKHG-KGPYTR 1572  
 Db 2204 HLQKNKYSYQGWNR 2219  
 RESULT 8  
 US-10-418-861B-55  
 ; Sequence 55, Application US/10418861B  
 ; Publication No. US20040010131A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: da Silva, Ana Claudia Raseira  
 ; APPLICANT: Parah, Shaker Chuck  
 ; APPLICANT: Quaggio, Ronaldo Bento  
 ; APPLICANT: Reinach, Fernando de Castro  
 ; APPLICANT: Ferro, Jesus Aparecido  
 ; APPLICANT: De Oliveira, Julio Cezar Franco  
 ; APPLICANT: De Laia, Marcelo Luiz  
 ; APPLICANT: Setubal Joao C.  
 ; APPLICANT: Furlan, Luiz Roberto  
 ; TITLE OF INVENTION: Isolated Xanthomonas nucleic acid molecules, proteins encoded thereon  
 ; FILE OF INVENTION: uses thereof  
 ; FILE REFERENCE: FAPESP 205.1 US  
 ; CURRENT APPLICATION NUMBER: US/10/418,861B  
 ; CURRENT FILING DATE: 2003-04-17  
 ; PRIOR APPLICATION NUMBER: US 60/374,620  
 ; PRIOR FILING DATE: 2002-04-22  
 ; NUMBER OF SEQ ID NOS: 85  
 ; SEQ ID NO 55  
 ; LENGTH: 1510  
 ; TYPE: PRT  
 ; ORGANISM: Xanthomonas  
 ; FEATURE:  
 US-10-418-861B-55  
 Query Match 4.1%; Score 361; DB 15; Length 1510;  
 Best Local Similarity 21.2%; Pred. No. 5.4e-19;  
 Matches 304; Conservative 158; Mismatches 466; Indels 504; Gaps 75;  
 QY 144 VLTGFNNAFDLK-----VPKKLNPAAGHAIYIDWNFEATOPRLNRIYDLDG 191  
 Db 250 ILTRPNGNFYTKKSANGTWTDPDVRRETLSEVDNGTLTGWQVTATDDSKQF--DLDG 307  
 QY 192 HDIPLNLEYQGLIKTILTFPGQEGYRTELRLNRLNLSIHN-----F 236  
 Db 308 ---KLTGISY-----TDGQLTLTVAGRLQQLQVTRGRRLLPAYQADRI 350  
 QY 237 SLGNENPLTWSFGYTPIGKNGILGOWITSMAPQGLKETV---NYSNNQGHFPQS--- 290  
 Db 351 QVGLPDGMVLAYGYDSQARL-----RSVTLTAAGAVASAIAGYDYG--ARFPDALTT 402  
 QY 291 -----ANLPVLPYVTLMLKQVPGQAPAI-QAEYSY-----TSHNYVG--GSN 330  
 Db 403 HRDEQOVYASVYDAQQRVRSVHGDPFGKIDETAYSGNTSTVSNALGNVTRTGIS 462  
 QY 331 GIWNNKLDNLYGLMTEYNYGSTESR-----YKQEGHQIVRIERTYNNYHLLTSECKQ 385  
 Db 463 KLGAQKAVTAQVGLCPCTVGAFKSRSYDNGYPPQEVDFAGVATDWRNRLGLASKIEA 522  
 QY 386 QN---GYIOTTTAYVAILIHNFDSPQSFQPLPKTKETWRSAD-----NSYRSE 432  
 Db 523 DNSSGGOKETLQT-----DWHPS-FRVP-TDQRTYDASDILVARTSWIYNSRQGA 570  
 QY 433 ITETTFDESNGPLTKVKKDKTKTKIISSTHWEYYPAGEVDNCPPEPGYGFTRPVKKIIQ 492

Db 571 LTVSRTPSGGS-TRV-----TTQRYCEDSD-----TAAG--NC----- 601  
 QY 493 TYPDSEFKDDPEKFIQRYVSLIGSQSHVTLKIEERHYSATOLANSTLFOYNTDKSELGRL 552  
 Db 602 -----SLPG-----LLASDGAFTDIADTSYTY-----YL 627  
 QY 553 LKQTECT-----KGNGKTYSVVHKFT-YTKQDD-----TLQOSHAIITHDNFTIH 597  
 Db 628 ADDASCTASACPHRKGDNRKVNALCQTEYLAYDAGRPLSIKDTNGIVT--DYTYH 685  
 QY 598 -----RSQVRSRYTGRLFSDTKDVTOMSYDKLGRLLTRTLNSGTPYANTLTVDYDEL 651  
 Db 686 PRGWLTSKVRG-----ADASSEADRIIDYVPTG--LVRQVTPDGAFTFTVD-AA 737  
 QY 652 NNLODDNRPFPVITTTDVNGNQLRNEFDGAGRHYVSQCLKSDSDGDKFYTIHTQYDEQGR 711  
 Db 738 HRLTD-----ITDNAGNTVHTLNDAGNRVKEDTKDAAATLK--RTLSEVYNQLGQ 786  
 QY 712 HHT--STYSD-----YLTNGRQQTDPKAVHLSMSKSYDNWGOI-----ANTHS 753  
 Db 787 LKQATAASDPTDFAYDANGNATKVTALATATQSEYDPLNRLSHILODVAGIKATKFA 846  
 QY 754 YGVSEKIT--VDPITLTATKQLOSNNVQ-----TGKEVTVYTPSQOPIQITLFDGAGH 806  
 Db 847 YDALDNLTKVTDPKGLDITDYNGFGLVKLTSPDTG--VTSYT-----YDSAGN 894  
 QY 807 LQSCHTLTRDGDWRVKETDAIGCTIYOYDNVNRVQITLPDGTIVNRKYVAFSTDTLI 866  
 Db 895 RAT-----QTDARGNTTAYSDALNRLTKVITYPS-----SLDVTY 930  
 QY 867 T-DIRVNGISLGQOQTFDGLSELSTQSDGGRVWAYTYSAGNDQCFSTVITPDGOFIHYQYQ 925  
 Db 931 TYDVTQTACTSG-ETF-SIGRLTKVQGGGAIQTCYNRFGD-LVRKYQTSNGTALVRYD 987  
 QY 926 PEL-----DDAVLQVASNE--ITQOFSYNPVTGALLKAVAGQSULTPIYYPGRL 973  
 Db 988 YTVGQLRRMYTPDGAUVYVYRNAQGTTOGVTPAGGS--RQVLLGNA---TYYPFGA 1042  
 QY 974 KMENINDMKKSYLWLT-----RGLNGYVTLTGTIQTOKISRDTHGRVATQIKDS 1021  
 Db 1043 AGWTYNGRTILARQVLDYRPAQIOTRPGGLDVG-----GFPAGNLALTALTA 1092  
 QY 1022 SLKTT---LNVDDLNRHIGSOVTDLATGHLMTTTPVDFGLNREIGRKLCDSSGHTLDIQ 1078  
 Db 1093 GNTPEICLGYDALGRLTG--LTDGVITGVI-----DG----- 1123  
 QY 1079 SWLKTQQLANRIVKINGVLQRTBOYSYDSR-NRLNQYKCDGACBPTDKYGHISIVTQNTFY 1137  
 Db 1124 -----YSYDATGNRL-----STKVG--TATQITYT 1146  
 QY 1138 DIYGNITACHTTFADGTEDHATPKFANPTDPCOLTEVHTHPDMPDNIRLKYDKACRVIN 1197  
 Db 1147 -----PNDSHLSAV-----AGVARIYDATG---N 1168  
 QY 1198 ITDNHGNENTYTLGRLONGQ--GSVYCYDPLNRLVSOXTDLQCELYYRETMVJNEV 1255  
 Db 1169 TTAIGGTARQVYTDTSGRMTQARRAGAV-----TMNRYN 1203  
 QY 1256 RNGEMI-RLLRGTETIIAQORASKVL-----LTGTDSSQSVIL----- 1292  
 Db 1204 GKGEQVRRFLGTNTYTLFDEAGHNLGVDYDSNGAPKQQAIIWLDDLPVGLLANANKLHYTE 1263  
 QY 1293 -----TSD-----KQNLQBSAYSAKHKSTAND--ASILG--YNGERADPV 1331  
 Db 1264 PDHLGSPRVVIDPTDPROVAVMTSLKGEAFNGTAPNPDGDDGAAAVLDMRFPQGRFDDAS 1323  
 QY 1332 GVTHLNGYRSYDPTLRFHTPDSLSFGF-AGGINFYSYCLGDPINRSDPSG 1382  
 Db 1334 GLNQ--NVFRDYEAATGRYQSD---PIGIEGGISTSYSLSSFPVKYIDVLG 1370



```
; Sequence 76214, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76214
; LENGTH: 843
; TYPE: PRF
; ORGANISM: Salmonella typhi
; US-10-282-122A-76214

Query Match 4.0%; Score 353.5; DB 12; Length 843;
Best Local Similarity 22.8%; Pred. No. 7.8e-19;
Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

QY 542 YNTDKSELGR-----LLKQTECTGKNGKTYSVVHKFTYTYKQDDTLQOQSHSITHDNTI 597
Db 32 HHTDSGQYRLDNWLAERSLCVDSMGT-----RCHWDAQGLVTAY 74
QY 598 RSOVRSRYTGR-----LFSDDTDKDVQMSYDKLGRLLTTLNSGTPYANTLYDYE 650
Db 75 RDEAGQMTTFRWSDEERLLLGMTDAQGGKRYVYDRLGHL-----TETHD-P 120
QY 651 LNNLQDNRPFP---VITTTDVNGNQLRNEFD-----GAGRHYVQCL 689
Db 121 LGRVEQTMHPVHQPETEYDAAGVAMRYEYDERGNLQAVSDPLHQRTVYGYDRH-QQV 179
QY 690 K--DSDGDGKFTYHTHQYDEQ---RH-----HTSTYSDYLTNGRQOTDPKVLHLSMS 738
Db 180 RITDARGGDKYL-----QWEDQOLMRHTDCSSQATWFYDERLRLERVTDAES--NSTR 232
QY 739 KSYDNWQCIANTHSYGVSEKIVDPITLPAKLOQSNMNVOTGKEVITYTSPQPIQI 798
Db 233 YSYDGNCHLTVFMAFDGRTERYQPD-----AAGRLVKYTSFAG---QI 272
QY 799 TLFEAGHLQSCHTLRFDGDRVRKETAIGQCTIYQYDNYNRYVIQITLPDGTIVNRKYA 858
```

```
Db 273 TEWQ-----RDQGRVREQTATGRTAYEYDAYGRLTTLNENGESYRFRY- 319
QY 859 PFSTDTLTIDIRVNGISLGQQTFDGLSRLTQSOD--GGRVWAYTYSAGNDQCP----- 909
Db 320 -----DVLDRVTETQTFDGGSSRRAYGYNALNAVTAIVYGGERG 356
QY 910 -----STVITPDQFIHYOYOPELDDAVLQV-----ASNITQOF 944
Db 357 GEIRHGLERDAAGRLTAKITPE---TRTEYRYDAADRLLLEIRRRHRDAAEGEPEVI-RF 412
QY 945 SYNPTVGTALLKAVAEG-----QSLTPIYPSGRLLKMNINDMKMSYLVLTURGLEN 995
Db 413 SYDSAGNLLSEETAQVLQHRVDVQCNRTEQMPDGR-----TLRYLYY 456
QY 996 GYTDLTGTGIQIKISRDTHGRVTVQIKDSSIKTILNYVDLNRHIGSQVTDLTATGHMLTTTVEF 1055
Db 457 G-----SGHLOQI-----NLGRDVISEFTR----- 476
QY 1056 DGLNREIRGRKLCDDSSGHTLDIQQSWLKTQQLANRIV--KLNGVLQRT---EQSYDSRNR 1110
Db 477 DHLHREVQR-----SQGR-LDMRMVYDRGTGRLLTKCKMRGVVPETFDREYAYSGQDE 531
QY 1111 LNQYKCDGAECPDXYGHSIVTQNTFYDIYGNITAC-HTTFADGTEDHATKFNPTDPC 1169
Db 532 LLK-----KHSRQGVTDYF-YDTTGRTACRNEAYLDSQWYDAA---ANLLDRR 577
QY 1170 Q-----LTEVHHTHPMDPNIRLYKDKAGRVINITDNGHNTENFTYDTL 1213
Db 578 QETAQAGAGSVVPENRITSYRGLH-----VRYDEVGRVVEKGRGNG-TQHYRWDAE 628
QY 1214 GRL-----QNGQGSVYG--YDPLNRLVSQKTDLDCLLYYRETMLVNEVRNGEMIRLLR 1265
Db 629 HRLTEVAVTRGTGTVRRYGYVYDAPGRVKEHLEDAEGPKYNRITFLWDMGLAQECRLGR 688
QY 1266 TGETIIAQOR-----ASKVLLTGTD-SQOSVILTSDKQNLQSEA-YSAY 1307
Db 689 SSSLYIYSDRGSHEPLARVDRAPQEADEVLYYHTDVNGAPEMTDGGNIWEAGYQWV 748
QY 1308 GK--HKSTANDASILGYNGRADPVSGVTHLNGYRSDYPTLMRPHPTDLSLSPFG-AGG 1363
Db 749 GNLTHEKETRPVQQLRFQGVLDRETGLHY--NLRYFYDPDIGNFISGD---PIGLAGG 803
QY 1364 INPYSYCLGDPINRSDPSG 1382
Db 804 INLYQYA-PNPLSYIDPLG 821

RESULT 10
US-10-282-122A-47600
; Sequence 47600, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 47600
; LENGTH: 1515
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47600

Query Match      3.9%; Score 350; DB 12; Length 1515;
Best Local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

QY 481 YGTRFVKKIQTQYDSEFRDPEKFIQYRYSLLIGSQSHVTLKIEERHYSATQLLNSTLF 540
DB 543 FQCGFYAYDEHGMWMTQRTDQTDVRYRYDTAG-----RVVE-----TGTRO 586
QY 541 QYNTDKSLGRLKQTECKGKTSVVKHFTYTRQDDTLQOHS-----IITHDNFTI 596
DB 587 GYHT-----GRFIVEAGCT-----RVIDVDGEWYAYNDEGLVTAETDPLGCHTYSWEWL 636
QY 597 HRSQVR-----SRV--TGRIFS-----DTTKOIVT-----QMSY 624
DB 637 GRLMARIDPLGRDYDERGQLTSSVSSGRTVDFDYDEQSLTGARLPNGGTIKLEY 696
QY 625 DKLGRLTRTNSGTPVANTITYD-----EL-----NNLQDNRPFVITTDV-- 669
DB 697 DHLRLIART-----EPDGNKTTYGRGELLRVVQGDRETRLDYDR-----LRLTDIEL 748
QY 670 -NGNQLRNEFPDGRHVSQCLKSDGDKFTYHTQYDQGRHHTSYSDYLTNGR--- 725
DB 749 PTGARFRKIDALGR-----LLEETSPDG-----HVTRYD-----YADGPANPRGLL 790
QY 726 -CQTDPPDKVHLSMSKSYDNWQCIANTHWSYGVSEKITV---DPITLTATKQLQSNNSNVQ 781
DB 791 SAVTRPD-----GSVSRAYN-----SESLPVEWIDPLGRIQI-- 824
QY 782 TGCEVTTVTPSQPIQITLFDGAG-----LQSCHTUTRQGWDRVRK 823
DB 825 -----TYGPFID-LLTASIDAAGHATRFYDHDATRLTKVINALGETYTYRYDAAGSLAA 876
QY 824 ETDALGQCTIYQYNYNRFVIGITLPDGTIVNRKYPASTDTLITDIRVNGISLQO--QTF 881
DB 877 EIDWGRATEYDRDAVGRLLTKLPDG-----GQWRYTY 910
QY 882 DGLSRLTOSQDGRWYAYTYSAGNDQCFSTVITPDGQFIHYQVQPELDDAVLQVANSNEIT 941
DB 911 DASDRLEIDAGDKVLAVRYDASG-----RLASAEVQGEHHTVT 949
QY 942 QQFSNPNVTGALLKAVAGQSILPIYPSGRKMKMENINDMKMSYLATRLGLENGYTDLT 1001
DB 950 -RFAYDR-NGRLIGEDQHGELLRHVYDAGQRLR-RMTPRRETTYA----- 993
QY 1002 GTIQXISRDTHGRVTKIDSSIKITTLNYDDLLNRHIGSQVTDL-----ATG-----HMLT 1050
DB 994 -----DVSGALTQVG-----QLTIRSDGLGRIGREAGDFVAQOQYDALGIRRRQIAG 1041

```

```

QY 1051 TTVEPDGLNRREIGRKLCDSSGHTLIDIOQSWLKTOQLANRIVKLVGLQRTQEQSYSDSRNR 1110
DB 1042 PAVAFDALQADPARAL-----BOLTRQV-----YHYDAAGQ 1072
QY 1111 LNOYKCDGACPT---DKYGHISIV-----TQFTYDIYGNITACHTTADGTEDHATF 1160
DB 1073 LERVDTGADTLTYORDERGQII CAESLLQPSHFYDAVWNI-AAHQORA----- 1121
QY 1161 KFAPTDPCQLTEVHHTH---PMPONIRLKYDKAGRVINIT-DNHG---NTENFTYDT 1212
DB 1122 ---PVD-----AHYRGGLPQGVYARYKYDARGRTIEKTVQGVPRPKTWQYWDG 1171
QY 1213 LGRL-----ONGQSVYGYDPLNRLVSQKTTDCLLYRET-----MLVN---EV 1255
DB 1172 LNRLVVVTPBERGVWA--YRYDAFNRIEKO-----QVGRETVKFLWDGEMLAERWIEQ 1224
QY 1256 RNGEMIRLLRTGETI-----IAQRASKV---LLTGTDSQSQSVILTSKQNLQSQE 1302
DB 1225 RDG-----TTGQVVWTHIEPQSFLPLAQETDDGLFILTDOIGRPKTVDFEQGPVWKA 1278
QY 1303 AYSAYGK---HKSTANDA-----SILGVNGERADPVSGVTHLNGYRSYDPTLMRFH 1351
DB 1279 AYSLWGLLPVKRPANDACGATSIDTTLRFSGQWADDETGLNLYNARY---YDPDSQOYL 1336
QY 1352 TPDLSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
DB 1337 SAD-----PIGLLGARTQAY-VHDFSQWIDPLG 1364

RESULT 11
US-10-282-122A-43059
; Sequence 43059, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/40/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1

```

```
; SEQ ID NO 43059
; LENGTH: 1397
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43059

Query Match
  3.9%; Score 342; DB 12; Length 1397;
Best Local Similarity 21.4%; Pred. No. 1.6e-17;
Matches 256; Conservative 157; Mismatches 382; Indels 404; Gaps 63;

QY 418 KTW-ROADNSYRSEITETTFDESGNPLTKVI-----KDKTKQKIIIS--PS-----461
DB 218 RTQTFHRAAGFSGEITGV- DGAGRHFLVLVTAQRAEABEQQAISGGTSPAPDPT 276
QY 462 -----THWEYPPAGEVDCNCPPEP---YGF- -----RPVK 488
DB 277 LPTTEYGRDNGIRLSAVALTHDPEP-----ENLPAALVRYGWTGPELAAVYDRSNT 331
QY 489 KIIOTPYSEFK-----DDEPKFIQYRSLIGSOSHVTLKIEERHYSATQLLNSTL 539
DB 332 QVRSFTYDDKYGRMVAAHRTGRPE--ICRYD--SDGRVTEQLNPAGLSYT-----379
QY 540 FOYNTDKSELGELLKQTEC--TKGNGKTYSVVHKFTYTKQDDTLQOSHSHITHDNFTIH 597
DB 380 YQYEDRTITDLSNRREVLHTQEGG-LKRWVK-----EHADGSVT 421
QY 598 RSQVRSRYTGRFSDTDTKDIWTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657
DB 422 QSQPDA--VGRLAQTDAAAGRTTEYSPVVTGLTRIT--TPDGRASAFY-----468
QY 658 NRPFFVITTDVNGNOLRNEFDGAGRHVSQCCKSD-----GDGKFTYIHTQQVDEQGRH 712
DB 469 NHHSLTSATPDGLIRREYDEWGRGLIQTAPDGDITRYVDNPHSDLPFCATEDATGSR 528
QY 713 HTSYSDYLINGRQOTDPKVLHLSKSYDNMGQIANTHWSYGVSEKITVDPI--TLATK 771
DB 529 KTMWRSY---GQLSFDCSGYTRYDHRFGQVTAHREELSGVRAVDSRGQLLAVK 585
QY 772 QLOSNNVQTKGVTTTPSQPIQLTFDEAGHLSCHTLTRDGNDRVRKETDAIGQ--830
DB 586 DTQGHETRYE-----YNAAGDLAT--VIAPDG--SRNGTQYDAMGKA 623
QY 831 -CTI-----YOYDYNRVLIQITLPGTIVNRKYAPSTDLTIDIRVNGISLGQTF--881
DB 624 ICTTQGGITRSMEDVAGRVIRLSENGS-----HTFRYDVLDRLIQETFGFGRTQRYH 678
QY 882 -DGLSRLTQSDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQVQPELDDAVLQVASNEI 940
DB 679 HDLTGKLIRSEDEGLV-----THWYD--EADRLTHRTVNGET 714
QY 941 TQPSYNPVTGAL--LKAVAGCSLTPIY-YPS-GRUKMENI-----NDM---KMSY 986
DB 715 AERQWYDE-RGWLTDISHISEGHRVTYHYGYSKRLASEHLTVHHPQTNELLWQHETRH 773
QY 987 LWTLRGLENGYTDLTGTIQKISRDRTHGRVTQIKDSSIKTTLNVDLNRHIGSOVTDLATG 1046
DB 774 AYNAAGLAN--RCIPDSLPAAVEWLYTG-----SGWLSGKLG 808
QY 1047 HMLTTTVEF--DGLNRBIGRLCDSSGHTLIDQOSWLKTOQLANRIVKLVNGVLTQEQVS 1104
DB 809 D--TPLVEYTRDLRHRETLR-----SFRYELTAYTPAGQLOSO--HLNLSLLS-DRDYT 858
QY 1105 YDSNRNLNOYKCDGACPTDKYGHISIVTONFTYDIGNITACHTTAD-----GTEDHA 1158
DB 859 WNDGELIR-----ISSPRO-----TRSYSGTTRGLTGVTHTTAANLDIRPYTDDPA 906
QY 1159 TFKFANPTDPCQTEVHTHPD-----MPDN-----IRLKYDKAGRVINIID--1200
DB 907 GNRLPDP-----ELHPDGSALSMWPDNRARIADAHYLYRYDRHGRLTEKTDLIPGV 956
QY 1201 ---NHGNTENFTYDTLGRQLONGQGSYVG-----YDPLNRLVSKQKTDLDCHEL-----1244
DB 957 IRIDDERTHYHDSQRLVHYTRTQYAEPLVSRVLYDPLGRVRAKVRWRERDLTGWM 1016

; RESULT 12
US-10-282-122A-43060
; Sequence 43060, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43060
; LENGTH: 1411
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43060

Query Match
  3.8%; Score 340.5; DB 12; Length 1411;
Best Local Similarity 20.4%; Pred. No. 2.2e-17;
Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

QY 418 KTW-ROADNSYRSEITETTFDESGNPLTKVI-----KDKTKQKIIIS--PS-----461
```



779 NVOTGKEVT-TYTPSQOPIQITLFDGAGHLSCHTLTRDGDWRVVKETDAIGQCTIYOYD 837  
503 TSSRGETVRYDDAHSELPAITTDATG---STRQMTWSRYGQLLAFTDCSGVQTRYEYD 559  
838 NYNRVIOITLFDGHIYVNRKXAPSTDLITDIRVNGISLQCOQFDGLSLRJTQSQDG-GRV 896  
560 RFQM-----TAVHRE-----EGISL-YRRYDNRGRLTSVKDAQGRE 595  
897 WATYSAGNDQCPSTVITPDQPFHYOYQ----- 925  
596 TRYEYNAAGDL--TAVITPDGNSETQYDANGKAVSTTQGLTRSMYEDAGRVIISLTNE 653  
926 -----BELDDAVQVNASNEITQOFSYNPVTGALLKAVAGQSLTPIIYPSGRLKME 976  
654 NGSHSVFSDALDLVQGGFDGRTQRYHYD-LTGKLTQSEDEGLVILWYDESDBRTHR 712  
977 NINDMKKSVLWTLRGLENGUYDITLGTQIKISDTHGRVTOIKDSS-----IKTLNYYDDL 1032  
713 TVNGEPAEQOY-----DGHGWLTDISHLSEGRHVAVHYGYDDK 751  
1033 NRHIG--SQVTDLATGHL----- 1049  
752 GRLTGECQTVENPETGELLQWHEKHAYNEQGLANRVTPSLPPVEWLTGSGYLAMKML 811  
1050 --TTTVEF--DGLNREIGRLCDSSH--TLDIQOSWLKTOQLANR----- 1089  
812 GGTPLVEYTRDLHRETVRSGMAGSNAAYELTSTYTPAGLQSQHLSLVYDRDYGWS 871  
1090 ----IVKINGVLQRTQOYSYDSNRLNOYKCDGAC-----PTDKYGHSI----- 1130  
872 DNGDLVRISGPRQ--TREYGVSATGRLESVETLAPDLDIRPYATDPAGNRLDPPELHPDS 930  
1131 -----VTQN-----FTYDIYGNITACHTTFADG---TEDHATKFFANPTDPCQLTEV 1174  
931 TLTWVPDNRJAEADHYVRYHDEYGRLTETKTRIPAGVIRTDERT----- 975  
1175 HHTHPDMPDNIRLKYDKAGVINITD-NHCN---TENFTYDTLGR----- 1215  
976 HHVH-----YQSCHRLVYTRIOHGEPLVESRYLYDPLGRMAKVRWRERDLTG 1025  
1216 ---LQNGQSVYGYDPLNRLVSKQTTDLDCELYR----- 1247  
1026 WMSLSRKPEVTWYWGDD-DRLATTVQDTTTRIQTVPYSPGFTPLIRVETENGEREKAQRS 1084  
1248 --ETMLVNEVRNG-----EMIRLL-RGETITIA-----QQASKV 1279  
1085 LAETLQOEGSENGHGVVFPABLVLVLDLLEEEIRADRVSESRANLAQCLTVEQLARQV 1144  
1280 LLTGDSQOS-----VILTSQKONLSOE-YSAYGKHKSTANDASILGYN----- 1323  
1145 EPEYTPARKAHLXCHDRGLPLALISEDGNTAWASAYDEWGNQLEENPHHV--VQPYEL 1202  
1324 -GERADPVGSVTHLNGVYSYDPTLMRFTPDLSLSPFG-AGGINPYSYCLGDPINRSDPS 1381  
1203 PGQCHDEESGLY--NRHRYDPLQORYITQD---PMGLKGGWNLQYPL-NPLOCIDPM 1256  
1382 GHLSWAQWTGIGMGIAGLLTIATGMAIAAGGIAAAIASTTALAFGALSVTSDITS 1441  
1257 GLL--QWDDARS-----ACTGV-----CGVLSRIIGPSKFSSTADALD----- 1296  
1442 TVSGALEDASPKASSILGWVSMGMGAAGLAESAIGKGTKLATHGLGAFEDGENALLKSTS 1501  
1297 ----ALKETQNRS-----LCNDMEYSYGIVCKDTNG-----KYFASKAETDNLK--K 1336  
1502 ESSRIKXGVTSLDRIVNEEGOVIKDHSRGVTDNFMGKGCQAILVHGDK--DGFLYHT 1559  
1337 ESYPLRKCPGTGDRVAATHHG---ADSHGDYVDEFFSSDKNLVRSKNNLEAFYLAT 1393  
1560 EGNKH---NGKGPY 1570  
1394 PDGRFEALNNKGEY 1407

RESULT 14  
US-10-282-122A-42617  
; Sequence 42617, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Cart, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 42617  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-10-282-122A-42617

Query Match 3.8%; Score 339; DB 12; Length 1426;  
Best Local Similarity 20.2%; Pred. No. 3e-17;  
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;  
QY 572 KFTVTKDDTLTQSHSITTHDNFTIHRSQVRSRYTGRLSFSDTDKDIVTQMSYDKLGRLL 631  
Db 317 RYTTTEAGELL-----AYDRSNTQVRAFTDAQHPGRVWYHRYAGREMYRYDDTKRVV 372  
QY 632 TRTNSGTPYANTLITYEYELNNLQDDNRPPFVITTD----- 668  
Db 373 EQLNPAG-----LSYRY-----LYEQDR-----ITVDSLNRREVLTHTGGAGLXRVVKE 418  
QY 669 -VNGNQLRNEFDGAGRHVSQCLKSDGDKGYTHITHTQOYD-----EQGHHTSTYSYL 721  
Db 419 LADGSVTRSGYDAAGRLTAQ--TDAAGRTEYGLNVVSGDITDITTPDGRKTEKYND-- 474  
QY 722 TNGRQOT---DPDKVHLSMSKSYDNWQCIANTHWSYGVSEKIVDPITLTATKQLQNSN 778  
Db 475 --GNQLTAVVSPD--GLESREYDEPCRLV-----SE 502  
QY 779 NVOTGKEVT-TYTPSQOPIQITLFDGAGHLSCHTLTRDGDWRVVKETDAIGQCTIYOYD 837  
Db 503 TSSRGETVRYDDAHSELPAITTDATG---STRQMTWSRYGQLLAFTDCSGVQTRYEYD 559

QY	838	NYNRVIOITLPDGTIVNRKYAPSTDTLITIRVNGISLQOQTFDGLSRLTSQDG-GRV	896
Db	560	RFQOM-----TAVHRE-----EGISL-YRRYDNRGLTSVKDAQGRE	595
QY	897	WAYTYAGNDQCSTVITPDGQFIHYOQ-----	925
Db	596	TRYEYNAAGDL--TAVITPDGNRSETQYDAWGKAVSTQGLTRSMXEYDAAGRVLSLTNE	653
QY	926	-----PELDDAVLOVASNEITQOFSYNPVTGALLKAVAEQGSULTIYYPGSRKME	976
Db	654	NGSHSVPSYDALDELVOQGFDCFTQRYHYD-LTGKLTQSEDEGLVILWYDESDRITHR	712
QY	977	NINDMKMSYLWTLRGLENGELYDTLGTQIKISDPTHGRVTOIKDSS-----IKTILNYDDL	1032
Db	713	TVNGEPAEQWQY-----DGHGWLTDLSHLSGHRVAVHYGYDDK	751
QY	1033	NRHIG--SQVTDLATGHML-----	1049
Db	752	GRUTGECQTVENPETGELLAQHETKHAYNEOGLANRVTPDSLPPVEWLTVGSGYLAGMKL	811
QY	1050	--TTTVEF--DGLNREIGRKLCDSSGH--TLDTCQSWLTKTOQLANR-----	1089
Db	812	GGTPLVEYTRDLRHRETVRSFGSMAGSNAAYELTSTVPAGOQSQOHSLSNVYDRDYGWS	871
QY	1090	-----IVKLVGLVORTEOYSVDSNRNLNQYKCDGAEC-----PTDKYGHSI-----	1130
Db	872	DNGDLVRISGPRO-TREYGYSAUGRLSVETLAPDLDIRPYATDPAGNRLPDPELHPDS	930
QY	1131	-----VTQN-----FTYDIYGNITPACHTTFADG--TEDHATKFPANPTOPCQLTEV	1174
Db	931	TLTWPNDRNIAEDAHVYVRHDEYGRUTEKTRIPAGVIRTDDBERT-----	975
QY	1175	HHHPDMPDNIRUKYDKAGVINITD-NHGN--TENFTYDTLGR-----	1215
Db	976	HHYH-----YDSQHLRVFYTRQGEPLVESRYLDPLORMKAVWRERDLTG	1025
QY	1216	-----LQNGQSVYGYDPLNRLVSOKTDTLDCELYR-----	1247
Db	1026	WMSLSRKPEVTWYGWDG-DRLLTQTQDTTRIQTWEPGSPFLIRVETENGEREKAQRRS	1084
QY	1248	-----ETMLVNEVRNG-----EMLELL-RTGETIIA-----QORASKV	1279
Db	1085	LAETLOQEGSENGHGVVPAPELVRLDLRLEEIRADRVSSESRAWLAQOGLTVEQLARQV	1144
QY	1280	LLTQDTSQQS-----VILTSDKQNLQSOBA-YSAYKHKSTANDASILGYN-----	1323
Db	1145	EPEYTPARKAHLXCHDRGLPLALISEDGNTAWSAEYDEWGNQNLNEENPHHV--YQPYRL	1202
QY	1324	-GERADPVSGVTHLNGYRSVDPTLMFHPFDLSLSPFG-AGGINPYSYCLGDPINRSDPS	1381
Db	1203	PQGOHDEESGLIY--NRHRYVDYPOGRIYITQD---PMGLKGWMLYQYPL-NPLQOQIDPM	1256
QY	1382	GHLQWQAWTGICMGIGIAGLLLTATGCGMAIAAAGGIAAAIATASTSTTALFAGLSVTSITS	1441
Db	1257	GLL--QIWDDARSG-----ACTGGV-----CGVLSRIIGPSKFDSTADAALD-----	1296
QY	1442	IVSGALEDASPASIIIGWSMGGAAGLAESAATKGGTKLATHLGAEDCENALIAKSTS	1501
Db	1297	-----ALKETQNRS-----LCNDMEYSGIVCKDTNG-----KYFASKAETDNLK--K	1336
QY	1502	BSSRIKMGVTRSLDREIVRNEEGQVVKDHSRGYDNFMGKGEOAILVHGDK--DGLFYHT	1559
Db	1337	ESYPLKRKCPTGDRVAAYHTHG--ADSHGDYVDEFFSSSDKKNLVRSKONNLEAFYLAT	1393
QY	1560	EGNKH---NGKGPY	1570
Db	1394	PDGRFEALNNKGEY	1407

RESULT 15

US-09-815-242-10384

US-09-813-242-10384  
: Sequence 10384, Application US/09815242







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:44:02 ; Search time 25 Seconds  
(without alignments)  
6437.137 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879  
Sequence: 1 VYIKFLKLFRRITWSDNNEF.....PRKIILGRTEKTVKPKTRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*\*

1: Pir1:\*\*

2: Pir2:\*\*

3: Pir3:\*\*

4: Pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1214	13.7	709	2 S38241	hypothetical prote
2	912.5	10.3	528	2 S38242	hypothetical prote
3	432	4.9	2334	2 S32920	cell wall-associat
4	429.5	4.8	336	2 S38239	hypothetical prote
5	386	4.3	2167	2 AF1489	cell wall-associat
6	381	4.3	2183	2 T37218	hypothetical prote
7	365	4.1	1400	2 E90886	RhsE core protein
8	353.5	4.0	843	2 AB0339	Rhs-family protein
9	342	3.9	1397	2 C64805	RhsC protein precu
10	341.5	3.8	1404	2 E90658	RhsG core protein
11	341	3.8	1394	2 H91236	RhsH core protein
12	340.5	3.8	1411	2 E65145	RhsB protein precu
13	339	3.8	1426	2 H64780	RhsD protein precu
14	336.5	3.8	1404	2 E85509	hypothetical prote
15	331.5	3.7	1377	2 C65159	RhsA protein precu
16	331	3.7	1409	2 F91187	RhsA core protein
17	326	3.7	1399	2 A99120	RhsC core protein
18	322.5	3.6	1377	2 E86034	RhsA protein in rh
19	319	3.6	1397	2 A85570	RhsC protein in rh
20	316	3.6	1398	2 B85549	hypothetical prote
21	313	3.5	1398	2 H90698	RhsD core protein
22	297	3.3	1512	2 AH0439	probable membrane
23	280	3.2	1317	2 F83310	conserved hypoteth
24	278	3.1	985	2 B86084	hypothetical prote
25	276.5	3.1	1438	2 A10093	conserved hypoteth
26	271.5	3.1	1354	2 AG0538	Rhs-family protein
27	266.5	3.0	2515	2 S47008	tenascin-like prot
28	248.5	2.8	1616	2 E90704	Rhs core protein w
29	245.5	2.8	1645	2 H85554	hypothetical prote

30 245 2.8 2825 2 T14271 Doc4 protein, stre  
31 230 2.6 794 2 T36972 probable membrane  
32 230 2.6 849 2 T46253 hypothetical prote  
33 220.5 2.5 1185 2 A42404 collagen adhesin -  
34 219.5 2.5 1962 2 A32634 lactocepin (EC 3.4  
35 218.5 2.5 2406 2 A54148 odz protein - frul  
36 217 2.4 356 2 T37136  
37 216.5 2.4 2894 2 C64474  
38 215.5 2.4 2044 2 AB1180  
39 213 2.4 1487 2 AG2560  
40 211 2.4 3283 2 AC1018  
41 208 2.3 2703 2 H81193  
42 207 2.3 2893 2 A64556  
43 206 2.3 1902 2 S06997  
44 203.5 2.3 1386 2 S57664  
45 202.5 2.3 1349 2 A11476

## ALIGNMENTS

### RESULT 1

S38241

hypothetical protein - Coxiella burnetii

C;Species: Coxiella burnetii

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999

C;Accession: S38241

R;Thiele, D.; Willems, H.; Haas, M.; Krauss, H.

submitted to the EMBL Data Library, October 1993

A;Reference number: S38215

A;Accession: S38241

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-709 <THI>

A;Cross-references: EMBL:X75356; NID:G407370; PIDN:CAA53129.1; PID:G407397

Query Match

Best Local Similarity 13.7%; Score 1214; DB 2; Length 709;

Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY 554 KQTECTKGNGKTSYVHKFTYTKQDDTLQQSHSTTHDNFTIHRQVRSRYTGRLFSDT 613  
DB 3 RRAEVLTSKEGKKYQNTFAISQAEHLQKIDFTGGDGKITSISREQSRYSGLLSST 62  
QY 614 DTKDIVTQMSYDKLGRLLRLTLN-SGTPYANLTLYDYELNNLQDDNRPFFVITTDVNGN 672  
DB 63 DELGNVTQVEYDELGRLLTQTVNASSTTYASTTRYSYSLTDARGKVTAKTTVDPKGN 122  
QY 673 QLRNEFDGAGRHVSQCLKSDSG-----DGKPYTHIQOYDECGRRHTSTYSYDLTNGRQQ 727  
DB 123 QLRTYDGLGRNLKQERLDKDAVSKTGTWYTHQOQYDALGRESKITITQDVL-----R 177  
QY 728 TDPDKVH----LSMSK--SYDNWGQIANTHSYGVSEKITVDPIITLTKOLQSNNVQ 781  
DB 178 LDSEVGHAGSVLSISKWHDYSGWQNHLLTVFSDGYQERSVYDPIITRAIT--LQESGSQK 235  
QY 782 TGKVVTPSPQPIQITLFDAGHLQSCHTLTRDGMWRVREKETDAIGCQTIYQYDNYNR 841  
DB 236 LGQQLTEYNLAGLPIKVTQYDSQGTQESAHYVDGLGQLRKETDELQIITLYEYDHGR 295  
QY 842 VIQITLPGTIVNRKYAPFSDTLITIRVNGISLGQQTDELGLRLTQSQDGGRWAYTY 901  
DB 296 VTQITLPGTIIQSYAPHSTASLITGISVNNFMNGQTFDSLRLTETTSGGTSAPSY 355  
QY 902 SAGNDQCPSTVITPDGQFIHYQOPELDAVLQVNASNEITQQFSYFNVPYTGALLKAV-AEG 960  
DB 356 ENAS-SVPAAVTAPTGETSVSEYKELGNVKKISAPEILQTWYDALTGAMTSATQAG 414  
QY 961 QSLTFIYPSGRKME-NIND---MKMSVLTWLRGLNGHYTLDTGTGIQKISRTHGRVT 1016  
DB 415 MIQMTYPSGLLKNMTPDGAQAKSTAYTISLAGAPQSYTDFVGTQYDYDEHGRI 474  
QY 1017 QIKDSSIKTTILNYDDLNRHIGSQVTDLATGTHMLTTTTFEFDGLNREIGRKLCDSSGHTLDI 1076

Db 475 GIEDNKLKSLDADGRFTKQATDKTKGAVLSLTLYDDLNREIKREISASQSVLVI 534  
 QY 1077 QQSWLTKQQLANRIVKLVNGVORTEQYSYDSNRNINQYKDCGAECPDTKYGHSIVTQNT 1136  
 Db 535 ECTYQRNHLKERITQGRGRTTLRKEMFAYDSNRNLIETTCNGEARPDQPYKAIHQRTFS 594  
 QY 1137 YDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVI 1196  
 Db 595 YDAGNWKTKQTDPSG-RNTATYIY-SALIDTQLKNNHSDYDPKXITLEYDKAGMI 652  
 QY 1197 NITDNHGNTEFYDTLGRQ-NG---QGSYGVYDPLNRLVSQ 1235  
 Db 653 R--DEAGRT--LRYDALGRLOQVNGAGARGQYAYDVLNLTLSQ 692

RESULT 2  
 S38242  
 hypothetical protein - Coxiella burnetii  
 C:Species: Coxiella burnetii  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Nov-1999  
 C:Accession: S38242  
 R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.  
 submitted to the EMBL Data Library, October 1993  
 A:Reference number: S38215  
 A:Accession: S38242  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-528 <THI>  
 A:Cross-references: EMBL:X75356; NID:9407370; PIDN:CAAS3130.1; PID:9407398  
 C:Superfamily: Coxiella burnetii hypothetical protein

Query Match 10.3%; Score 912.5; DB 2; Length 528;  
 Best Local Similarity 40.8%; Pred. No. 6, 9e-43;  
 Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

QY 18 NEF-FTQANFTSAVSGVDPRGTGLYNIQTILGHIVNGNIGLPLPLTSLVSPLNKTDI- 75  
 Db 2 NELPYQATNFTISAVOGVDPRGTGLFTVMVLAELTGNLGNLDPFLFTLNYSLHLSNIC 61  
 QY 76 GFGIGNFGSLVYDRNLSLSSTGENYKVTETDKVKLOOKLDNIRFEDKLENCYRI 135  
 Db 62 GFGIGCSGISYDKNNKLLILSSERYKTEDWDSVYVROKKINPKFEK--IKNGYII 119  
 QY 136 IHKSGDIEVLG--FNNAFADLVKPKLLNPAHAIYIDMNFPEATQPLRNRIYDLDLGDH 193  
 Db 120 KYKNGKTELYNKKYGDNLF--LPQKIFSLGWLPLKLTWENRGQYVNLTKIEDAKD--- 173  
 QY 194 IPELNLEYOGLIKTILTLFPQKQEGVTELRFLNRLNLSIHNFSLGNENPLTWFSGYTPI 253  
 Db 174 -VLCKIDYQPSDWARITFPFGKTESYTPQLDFVNEYLWVTKSTSR--LWVSFNYDDV 230  
 QY 254 GKNGILGQMITSWAPGGLUKETVYNNYNNQGHHPQSANLPLVPLVTLMQVPGAGQPAI 313  
 Db 231 GAGNFT---LTQKSPSTGLTETWYQAGV--RFPDESCKPALPSVYNYRQSPGMGQFDI 285  
 QY 314 QAEYSYTSNHYVGGSN--GINNKLNDLYG-LMTEYNGSTGRRYKDKGHEHQDQVRIE 370  
 Db 286 VXEYEYTVSNLYGYGASLGKAWNEEDNLYNMDDIYYSSTEXKLIVDNR---LVSIS 341  
 QY 371 RYNNYHLTSCQKQNGYIQTETAYVAILGNHFDSPQSPQFOLPKYKTETWR--SADNSY 429  
 Db 342 RYNSYLLISITTRQNSCEVITDYAKPCLSPDKQPKQFOLPKKEKTKWRENSKQC 401  
 QY 430 RSEITETTESGNNPLTKVKKTKQKIISPSTHWEYPPAGEVDN---CPPEPYFTR 485  
 Db 402 RSEITTTTDPGNNLTKEPD-----GKTEYIYDYSKGETDKGIVLCPPEPFGVR 454  
 QY 486 FVKKIQTQYDSEF-----KDDPEKFIQRYSLISQSHVTLKIBERHYSATQLLNSTLQ 541  
 Db 455 FVKTQIVTPADSEFYAPVQQTITAYAQYPCITAGSLSYAVLQTOETLCSDDVL-----LLT 510  
 QY 542 YNTD 545

Db 511 INTD 514  
 RESULT 3  
 S32920  
 cell wall-associated protein precursor wapa (similarity) - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 21-Jul-2000  
 C:Accession: S32920; E69730; T47101  
 R:Foster, S.J.  
 Mol. Microbiol. 8, 299-310, 1993  
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis  
 rotein.  
 A:Reference number: S32919; MUID:93302506; PMID:8316082  
 A:Accession: S32920  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <POS>  
 A:Cross-references: GB:L05634; NID:9304177; PIDN:AAA22883.1; PID:g304179  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emmergon, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningstein, G.; Kroch, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A.; Others: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: E69730  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <KUN>  
 A:Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469  
 A:Experimental source: strain 168  
 R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.  
 Microbiology 141, 337-343, 1995  
 A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai  
 A:Reference number: Z24350; MUID:95219088; PMID:7704263  
 A:Accession: T47101  
 A:Status: preliminary; translated from GB/EMBL/DDB  
 A:Molecule type: DNA  
 A:Residues: 1-2334 <YOS>  
 A:Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782  
 A:Experimental source: strain BGSC1A  
 C:Genetics:  
 A:Gene: wapa; N17G  
 C:Superfamily: cell wall-associated protein wapa

Query Match 4.9%; Score 432; DB 2; Length 2334;  
 Best Local Similarity 20.4%; Pred. No. 4, 1e-15;  
 Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;

QY 44 IQITLGHIVG-NGNL-----GPTLPLTSLVSPLNKTDIGFGIGENFGLSVYDR 90  
 Db 896 IDIFSGQLNGATGNVIVNEEDLSIDRGPGGLGSRTRYNSLSDSHLFGQGW-----YADA 950  
 QY 91 KNSLLS:STGSENYKVIETDKTVKLOQKLDN-----LRFKDLKENCYRIIHKSGDIEVL 145  
 Db 951 ETSVISTDGGAMY--IDEDATTHRTKKADGTQPTGVVLELTETADQFIKTKD----- 1004  
 QY 146 TGFNNNAFDLVKPKLL-----NPAGHAIYIDWN-----FEATQPLRNRIYDLDLGH-- 192  
 Db 1005 ---QTNAYFPNKKGKLOKQVVDGHNNAVTVYNDKNQITAITASGRKLTFTYDE-NGHWT 1060  
 QY 193 -----DIPLNLEY--QGLI-----KTYTLTFPG 214

Db 1061 SITGPKKKVTVSYENDLLKKVTDIDGTVTSDYDSEGLRVKQYSANSTAKPVFTTEY-- 1118  
QY 215 QKEGVRTELRLNQLNHSNFS-----LGNENPLTWGFGYTPPIGKNGILGOWITSMT 267  
Db 1119 QYSGHRLKALNAKKEVTVSYDADKKTLLMQPNGRKVQVGYNEAGNP-----IQVID 1172  
QY 268 APGGIKETVN--YNNN-----QGHFFPQSANLPVLPIYTLMMKQVPGAG 309  
Db 1173 DAEGELKITNTYKGNVVDVNDVGTGKATESYQYDKGN-----VTSVKDAYGT- 1225  
QY 310 QPAIQAEYSYTHNYVGGSGNGIWNKLDNLYGLMTEYNYGSTEYRRYKKEGHDQIVRI 369  
Db 1226 -----ETVEYNNKNDV-----TKMKDEGNTVTDIADGLDAVSETDQSGKSSAAV 1271  
QY 370 ERTYNNYHLLTSECKQNGYIQTETAYAILGHNFDSQSQFOLPKTKTETWRSADNSY 429  
Db 1272 YDKYGNQIQSGKOLASNTILK-----DGSFPAQSGMNLTASKD----- 1311  
QY 430 RSEITETTFDSGNEPLTKVINKKTKQIIISPS-----THWEYPPAGVDNCPPEPYGFT 485  
Db 1312 RKKIS-VIADKSG-----VLGSKALEVLQSSTAGTDHGYSSATQVLEPNVITLSTG 1365  
QY 486 FVKKII---QTPYSEFKDPEKFIQY---RYSLIGSQSHVTIKIEBHYSATQLLNS-- 537  
Db 1366 KIKTDLAKRAYFENIDLRDKDKRIQWIHNEYSALAGKNDWT-----KRQITFTPANAG 1421  
QY 538 ---TLFOYNTDKSELGR-LKQTECTKGNGKTVSVVHKETYTKQDDTLQOS-HSITTHD 592  
Db 1422 AVVYMEVDHDKDKGKAWDEVOLEGEVSSSTPNVQNSFTSATENMVNSGASVDSEE 1481  
QY 593 NFTIHRQVRSRYTGRFLPSDTDKDIVTQMSYDKLGLLRT-----L 635  
Db 1482 GFNDVSLKAARTSASQAGSVTKQTQVVLGQSANDKPVVLTITGMSKASSVKFTDKDYSL 1541  
QY 636 NSGTPPYANTITYDEL---NNLQDNRPPFVI----- 664  
Db 1542 QANVTYAGSGTGIYNAKFPSTQEWNRAAVVIPTKPIKVDISILFQKSATGTWFPDDI 1601  
QY 665 -----TTTVDVNGNQLRNFDDGAGRHVSQCLKSDGDKGFYTIHTQYDEQGRHHT 714  
Db 1602 RLIEGSLTKSTYDSNGVYVKEDELGYATS---TDYDTGK---KTSSETDAKEKT 1654  
QY 715 STY---SDYLNGRQOTDPDKVHLSMSKSYNNWG-QTANT-----HWSYGVSESKI-- 760  
Db 1655 YTDQADQLNTMLNLSGTSILH-----SYDKEGNEVSKITRAGADQTYKFEYDVMGKLVK 1709  
QY 761 TVDPIITLATAQLQSNVNVOT---GKEVTTTTPSQOPIQITLFEAGHLQSCHLTLD 816  
Db 1710 TTDPLGNVLASEVDANSNLTKTISPNGEV-----SLSYD 1744  
QY 817 GWRVRKETDAIGCTTIQYDYNRVVIQITPLDGTIVNRKYAPFSTDLITDIRVNGISL 876  
Db 1745 GTDRVKSXSYNGTEKYIFTYDKNGN-----ETSVNVKEQN-----TT 1781  
QY 877 GQOTEGSLRLTQSQDGRVWATYAGNDOCPSTVITPDQFIHYQYQPELDDAVLQVA 936  
Db 1782 KKKIFNKNRLIELTRGGGQWTYPSDSKLKTF-----SWIH-----G 1821  
QY 937 SNEITQFSPYPTVGTALLKAVAREQSULTPTIYPSGRLKMNINDMKMSYLMWTLRLENG 996  
Db 1822 DQKGTNQFTN-----KLDQMIEMKDSSTSSYSFYDEN- 1854  
QY 997 YDLTGTIQKISDTHGRVTOIKDSSIKITLNVDDLNHRHIGSQVTLATGHMLTTTVEPD 1056  
Db 1855 -----GNVQ-----TFTGNGGGTFSFYDERNLVSSLHIGDKNGDILTESYEY- 1898  
QY 1057 GLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLVGLQRTQEDYVSDSRNRLNOYKC 1116  
Db 1899 -----DANGRTTINS-----ASKV-----QYEYKLNQLVK----- 1927  
QY 1117 DGACETDKYGHIVTQNTFYDIYIGNITACHTTFADGTED--HATPKFANPTDPCQLTEV 1174  
Db 1928 -----ETHEDGTVI---EVTYDGFGRKTV-TTIKDGSSTKTVNASPNMN-----QLTKV 1973

QY 1175 H-----HTHP-DWPDNI-----RLKYDKAGRINUITDNHG 1203  
Db 1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLATVATKKGEKPPATYKYDEKGNRIQKTVN-G 2032  
QY 1204 NTENFTYDGLRQLONGQSGVYDPLNRLVSOKTTDTLDCELY-YRETMLVNEVRNEMIR 1262  
Db 2033 KVTNYFYDG-----DSLNLVYETDADNNVTKSYTYGD-----SQLLS 2070  
QY 1263 LLRTGTETIAQORASKVLLTGTDSQSVLTSDQNLSQEAYSAYGK-HKSTANDA---S 1318  
Db 2071 YTEGKKFYHYHNAHDIITAIISDSTGKTV-----AKYQDANGNPTKTEASDEVKDN 2122  
QY 1319 ILGNGERADPVSGVTHLNGVRSYDPTLMRPH--PDSLSFPFGAGGPNPYCYCLGDPIN 1376  
Db 2123 RYRYAGYQYDEETGLYLMARY--YEPRNGVFLSLDPPGSDCDSLDQNGYAYGNPNPM 2180  
QY 1377 RSDPSCHLSQAWTGIWGIAGLLLTATGGMAIAAAG---GIAAAATASITTTALAGA 1432  
Db 2181 NVDPDGH--W-VLVVNAAGFA-----AYDGYKAYKSGKMGKGAWAASNFGPKIFRG 2231  
QY 1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGV---GAAGLAESAIKGGTKLATHLGA 1488  
Db 2232 ASRAYKFTK-----KAVKITGTHRHGLNQSIGRNG-----GRGVNLRALANA- 2273  
QY 1489 AEDGENALLKSTSESRKMGVTRSLDRB--IVRNEGOVVKDHSRGYTDNPMGKGEQAI 1546  
Db 2274 -----VRSPKVIKQPNGATKYVGKATVVLNKRKGVITAYG-----SS 2312  
QY 1547 LVHGDQDGFVHYTEGKN 1563  
Db 2313 RAKGSKHVHTHGKGNK 2329

## RESULT 4

S38239  
hypothetical protein - Coccidiella burnetii

C:Species: Coccidiella burnetii

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Oct-1999

C:Accession: S38239

R:Thiele, D.; Willems, H.; Haas, M.; Krauss, H.

submitted to the EMBL Data Library, October 1993

A:Reference number: S38215

A:Accession: S38239

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 &lt;TH&gt;

A:Cross-references: EMBL:X75356; NID:g407370; PIDN:CAA53127.1; PID:g407395

## Query Match

Best Local Similarity 4.8%; Score 429.5; DB 2; Length 336;

Matches 121; Conservative 48; Mismatches 91; Indels 93; Gaps 13;

QY 1348 MRFHTPDSLPFGAGGPNPYCYCLGDPINRSDPSGHLISQAWTIGMGIAGLLTIATGG 1407  
Db 1 MRINCPDSWSPFGAGGPNPYACDGPINRVDPNGLHSQAEIDIGLVGLVAVFTAG 60

QY 1408 MAIAAGGIAAAIAASTTATAFALSVTSDITSVSGALEDASPKASSILGWVSMGMA 1467  
Db 61 TSIAAAGIAISAISSASISLVVGTGLVAADVASIASGALEDANPQASATLGLISLUGG 120

QY 1468 AGLAE---SAIKGGTKATHLGAFAEDG-----ENALLKSTSESSRIKMGWT--RS 1513  
Db 121 PGAVSGLATAARAGKKL---ISGLAKGGKIRSQSPVQGI-SYRSLSRGDLPLRGPPHPQS 177

QY 1514 LDREIVRNEE-----GOVVKDHSRGYTDNF-----MGK 1541  
Db 178 LSRVTVAPEMRPAGLVNWHKVSQKSLGLVGHVFCADREIFGYEIRPIEFRRRRPITK 237

QY 1542 GEQAIL-----VHGDKDGFLYHTEG-----NKHNGKGPYRHTPEQLVDY 1581  
Db 238 RDIVILSGTHGRVHGDN---WTSQGLRRRPDILERAFFIEDVQVYKG-----QLNGR 285



DB. 2102 VILGFGKL 2110

RESULT 6  
T37218  
hypothetical protein SC2H4.02 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
R:Accession: T37218  
C:Oliver, K.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z21615  
A:Accession: T37218  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2183 <OLI>  
A:Cross-references: EMBL:AL031514; PIDN:CAA20596.1; GSPDB:GN00070; SCOEDB:SC2H4.02  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC2H4.02

Query Match 4.3%; Score 381; DB 2; Length 2183;  
Best Local Similarity 19.8%; Pred.No. 2.5e-12;  
Matches 366; Conservative 221; Mismatches 669; Indels 594; Gaps 91;

	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB	QY	DB												
	99	TGENYKVIETDKTVKLGQKK-----LDMLFEKDLKENCYRIIHKSGDIEVLTF 148	620	TGDS-----TDQSLWLDDEIKHTGRGTDSLDAVKFSHVMPN--XVDGPSDDI-----666	149	NNNAFDLKVPKLLNPAGHAIIYDW-----NFETQPLNRIRYYDDLGDHDIDL 196	667	--LSFERPRLRITVSETGAOTIVDYLPADCVAGTWPKADENTKRCFPVYWSFYGOEPI 724	197	LNIYEGLIKITLFP-GOKEGYRTLEFLNQLSIHNSFLG-----NENPL-----244	725	LDMFQYPVSVSRVTTPLOGSEA-----VQHTVEYSGGAWHYNDPDLTPAKER 773	245	TWSEFGYTPGKNILGW-----ITSWTAPG---LKETVYNINNOQHHPQSANLPLV 296	774	TWS-----QWRGYGVKTHYTGSGGTRAKTVTVYLRLMGMDRVTLGGDGKPTD 820	297	PYTLMKVPGAGQPAI-----QAESYTSHNHYGGGSNGIWNKNLNLNYGLMTE 346	821	PDRRRKAESVGTAAAVTSDQLAGFORESVAYDGKVEYSTVNDPWSKXT-----871	347	YNYGSTESRRYKDKEGHDIQVRIERYNNYHLLTSECKQONGY--IQTTETAY--YAIIG 402	872	----ATQHSYADEAY--YVRVGAS---HARTRITSRLNPYDRVTRTGTGYDDYGMPV 921	403	HNFDSQPSOLFPKTKTETWRSADNS-----YRSEITETTFD---ESGNPL 445	922	SVEDLGDDSVTGDEKCTRTWARNDAAGLTALVSRTIYGRACSVYDSALDIPADAARP- 980	446	TXVWKDKTKQLISPSTHE--YYPAGEVDCNPPEPYGFTFRVKKIQTYPVDSFKDDP 503	981	GDVVSDKAT---AYDTTTSATQKPTKGDA-RWSGRAKYGGDDRPMLQMTAVTDY-DTL 1035	504	EKFQYRYSLIGSQSHVTLKIBERHYSATQLLNSTLFQYNTDKSELGRLLLKQTECTKEN 563	1036	GRPQVR---NTDDVITSKTEYQPPAAPLTSTVV-----N 1069	564	GKTYSVVHKFTYTKDDTLTQQSHSITHDNFIHRSQVRSRYTGBLFSDDTTKDITVOMS 623	1070	AK-----GHRTTTVKDFAL-----GADLKVTQANGKVTESA 1100	624	YKGLRELTRLTL-NSGTPYANTLTVDYELNQLQDNRPFPVITTTDVNGNQLRNEFFDAG 682	1101	YDLSGRVTSVWLPNRSRALGKTANYGVYS--VKSTSLEFWSSAT-LNG-----DGG 1150	683	RHVSQCLKSDCGGKFYIHTHQYDEQGHHHTISTYSVILTGR-----725	1151	YATTVEIYDS-----LIRTEQVOAPAQGRGVIACTLYDGRGLEVFATAQADIWDNTAAP 1203

QY	726	----	QOTDPDKVHLSMKSXYDNQWQIAN--THWSYGVSE-----	759
DB	1204	AGKIVQIDGGQAPROTDSVVDGMRGVTAKVTKSYGVTEWAVDTAYRGDLVLTGAPEGSSA	1263	
QY	759	-KITVDPTILTATQLOSNNNVOTGKEVITYTPSQOPIQITLPDEAGHLOSCHTLTRDG	817	
DB	1264	NAVTDAGFTRERRDVIAGTQPACTDWTTRYA-----FDAADROKSIHAHDRSA	1313	
QY	818	W-----	849	
DB	1314	WYTYDILFGRQSVSTDPKGTTVTTEYDALDRAVKSTDGRCVELLFEYDVLGRKTMQWSA	1373	
QY	850	GTIVNRKVPSTDTLI-----TDIRVNGISLQO-----QTDGSLRLLTQSO-----	891	
DB	1374	KTGAN--KLAWSFDTLAKGOODAVRYEGGETGRAVTKQVTRYDPLPKVTNNELTLPAND	1432	
QY	892	---DGRVNAVYTSAG-----NDQCPSTVITPDGQFIHQVOPELDDDAVLQVANSNEIT	941	
DB	1433	PLVAAGVPARLAFSTGYNLQGTQKQAAAPAVAGLSAETVSITY-----DGLGQVLTAKGT	1487	
QY	942	---QOFSYNPVTGALLKAVAEGQSITPIIYYPGSRUKMENINDMKMSYLVMTLRLGLENGY	997	
DB	1488	TGVLQQAAYSP-----LGLDRQMTLATDPTGAKKVYLNND-----YEAGTRRLTRSY	1534	
QY	998	T--DLTGTI---QKISRDRHGRVTOIKDSSIKITTLANYDILNRHIGSOVTDLATGHMLTTT	1052	
DB	1535	VTDVHGFMLQELKYQDDAGNITSVDA-----TTLG-----GFGKADHQC	1576	
QY	1053	VEFDGLNRREIGRKL-----CDSGHTLDIQOSWLKATQOLANRIVKLNGVLQRTQESYD	1106	
DB	1577	FTYDGG--HRLSEAWTETADCSGRT-----VAGLGGAAPVWTSYQYD	1619	
QY	1107	S---RNLNQYKCDGAECPTD--KYG-----HSI---VT-----QNTYDIYGNITACHYT	1149	
DB	1620	DSGIRSKQTEHRMSGDDVTVEYEGYAEQOPHALSATVTTGAENASVYDDETN-----	1672	
QY	1150	FADGTEDHATFKPANPTDPCQUTEVHTHPDMPDINRLKYDKAGRIVINITDNHGTENFT	1209	
DB	1673	-----TE---TRPGVREATQTLDNWAEGRLAGVSEPAAGCKPAT	1707	
QY	1210	YDTLGRLQNGQSVGYDPLNRLVSKOTDLDCE--LYIRETNVMEVR--NGEM-----I	1261	
DB	1708	-----GTAIVYDAGDGLLRRTTDTGETVLYLGTTEVHLKVSNGAALKSQA	1756	
QY	1262	RLLTGETIIAQRASKVLITGTDSQSVILTSXNLNLSOEAYSA-----YCK	1309	
DB	1757	RTYKASSAVIARTST---AGVSGTKLTFLAGDHHGTSGLAINDTLAFXRKWTPTPGA	1812	
QY	1310	HKSTANDA--SILGVNGERADPVSGVTHLNGYRSDPTPLMFHTPDS--LSPFPGAGGINP	1366	
DB	1813	PRGTASGAWPDRDGLFGKPADAAATGLTQL--GARQYDPTDGTGRFLSDPLLEPKDKNTLNG	1870	
QY	1367	YSYCLGDPINRSDPSGHLSSQAWTGIWGIAGLLTIAT--CGMAIAAAGGIAAAIASTS	1424	
DB	1871	YAVASNPVNSDPG-----TSDGLG--GLLGAIGAIGGVVGVAVIGAVGAAI--TA	1919	
QY	1425	TTALAPGALSVTSDTITSIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATH	1484	
DB	1920	VGSLG3GGG-----GGWGGGTAPTSSG--GW-----TQPLTKQWTPGATY	1956	
QY	1485	LGAFAEDGENALLKS-----TSBS-----SRIKKGWTRSLDREIYRNEE-----	1523	
DB	1957	-----NFTIKSWDILFPNDPPOSLEMLASMPDWGIVS--DPKAANRWETSRSLPFG	2005	
QY	1524	-----GOVIKHS--RGYTDNFMKGEOAILVHGDK--DGFLYHTEGNKHNG--KGPYTRHT	1574	
DB	2006	MLWGGGYPLRHHQDFRG-----GDAFTSILAQDETISGLRSKMVGQARKGTGKAFAKEV	2059	
QY	1575	PEOLVDYLKDNINVDLTQGGKPFVHLLSCYKSSGGAADKMAKYNRPVI-----	1623	
DB	2060	GFOQYVD-----EGPEGSPWYKFNLSRGAARDIAGVLTNAGVGTGNQADAPLG	2107	
QY	1624	AVSNKPTISQGLARIERKDFELKSTYHSYDPRKILIGRTEKTVKPKTRP	1673	

Db 2108 TYSKARIS---INKKESVTLKPSAWNGSD---WSATHVPRSNWP 2150  
 RESULT 7  
 E90886  
 Rhs core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain  
 C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C;Accession: E90886  
 R;Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: E90886  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1400 <HAV>  
 A;Cross-references: GB:BA000007; PIDN:BA035484.1; PID:gl3361527; GSPDB:GN00154  
 A;Experimental source: strain O157:H7, substrain RMD 0509952  
 C;Genetics:  
 A;Gene: ECE2061  
 C;Superfamily: rtsf protein  
 Query Match 4.1%; Score 365; DB 2; Length 1400;  
 Best Local Similarity 20.9%; Pred. No. 8.9e-12;  
 Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;  
 QY 418 KRTW-RSADNSVRSHEITTTTDESIGNLTKVI-----KDKTKIISFTWEY 466  
 Db 218 RLTYREAAAGLAGEITGV-DGAGREFRLVLTQAQAEARKQHTASLSPDT--- 272  
 QY 467 YPAGEVDNCPPEPGYGFTRFKKIIQTPYDSEFKDDPEKFIQYRSLIGSOSHVLKIEE 526  
 Db 273 -----RPLSDSAFFDTLPCTEY-----GPDGIR----- 297  
 QY 527 RHYSATQLNLSLTFQYNTDKSELGRLLKQTECTKENGKTVS-----VHKFTYTKOD 579  
 Db 298 --LSAVLWTHDPAYESLPGLARY-----TYTEAGELLAVDRNTQVRAFTDAQH 349  
 QY 580 DTLCQSHSITTHDNFTIHSQVRSRYTGRFLSFDTDKDIVTQMSYDKLGRLLTSLNSGT 639  
 Db 350 PGRVAH-----RYAGR-----PENRYRYDDTGRVVEQLNPAGL 383  
 QY 640 PYANTLTYDYELNLODDNRPFFVITTD-VNGNOLRNEFGAGRHVSQCLKSDGDKF 698  
 Db 384 SY-----RYQYE-----ITVTDLSNRREVLTGEGAG--LKRYYVKELADG-- 425  
 QY 699 YTIHQYDEQGRHHTSYSDYLTMGR--QOTDPPKVLHLSKSYDNWGQIANTHWSYGV 756  
 Db 426 -----SVTHSGYDAAGRLTAQTD-----AAGRTEYGL 453  
 QY 757 SEKITVDPILTATVQLQSNNSWOTKEVITYPSQOPIQITLFDAGHLQS-----CHT 812  
 Db 454 N--VWSGDITDITDPGRETKEFYNDGNOLTAUVSPDGLERRAYDEPGLVSTSRGSD 511  
 QY 813 LTRDGDVRKE-----TDAIGCCITQYDNNRVNIQITLDPGIVNARKYAPESDTLIT 867  
 Db 512 VIRAYDNPHELPAITTDATSTQMTWSRYGQLLAFTDCSGYQTRYEYDRFQMTAVH 571  
 QY 868 DIRVNGISLQOTFDGLSLTOSQG--GRWVAITYSAGNDCPSTVLTDPQPIHYQYQP 926  
 Db 572 --REGIS--RYRYDNRGLTSVKDAQGHETRYEYNAAGDL--TAVITPDGNSRETY-- 624  
 QY 927 ELDDA---VLOVASNEITQQFSYNPVTGALLKVAEGOSLPIIYPSGLKLMENINDMKK 983  
 Db 625 ---DAWGKAVSTQGGGLTRSEYD-LAGRITTLNENGRSEFTYDA----- 667  
 QY 984 MSYLTWLRLENGYDITCTQKISRSDTHGRVTKQSSIKTTLNYDNLNREHG----- 1037  
 Db 668 LDRLVQQRGFD-----GRTORHYVLTGKLTQSEDEGLVTLWHYDESRLTHRTVNGE 720

QY 1038 -----SOVTLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLDI-----QQS 1079  
 Db 721 PAEQWQYDEHGWLTEISHLSEGHQVAVHYGYDDKRLAGERQTVHNPETCELLQWHEH 780  
 QY 1080 WLKTOQLANRIVKLVLTQTEQSYSDS-----RNRL-----N 1112  
 Db 781 AVNEQGLANRVP--DSLFRVWLTGYSGYLAKMKGGTFLVETFRDLRHRETVRSGNN 838  
 QY 1113 QYKCDGAECPDKYKH-----SIV-----TONFTYDIYGNIT 1144  
 Db 839 AVELTSTVTPA---CHLQSRLNSQVYDRDYDNDNDGLVRISGPRQWEGYSATGRLE 895  
 QY 1145 ACHTTFADGTEDHATFRFANPTDPC--OLTEVHHTHPD-----MPONIRL-----KY 1189  
 Db 896 SVRTLASD-----LDRIPIYATDPAGRLPD--PELHPDSTLTAWPDN--RIAEAHVYVRH 948  
 QY 1190 DXAGRVINITD-----NHGNTENFTYDTLGR-----NQGQSV---YCYDPL-- 1229  
 Db 949 DEYGLTEKTRIPAGVIRTDDETHHYHYSQHLVFTYTRIQHGEPLVESRYLYDPLGR 1008  
 QY 1230 -----NRLVSQKTDITLDCELYR----- 1247  
 Db 1009 RMAKVRRRERDLTCWMSLSRKPEVTWYGDGDLRTVQDTTRIQTVYEPGFTPLIRV 1068  
 QY 1248 -----ETMLYNEVRNG-----EMIRLLTGETIIAQORAS----- 1277  
 Db 1069 ETENGEREKAQRSLAETLQOEGSENGHVFPFAELVRLDLRLLEEIRADRVSESRAWL 1128  
 QY 1278 -----KVLITGTSQ--QSVILTSKQNLQSEAYSAVGHKST 1313  
 Db 1129 AQCGLTVEQLARQVPEYTPARKVHFYCHDRGLPLALISEDNTAWRGYDEWQNLNE 1188  
 QY 1314 ANDASI---LGNGERADPVSGVTHLGNYSYDPTLMRFHTPDSLSPFG--AGGINPYSY 1369  
 Db 1189 ENPYVHLQPYRLPQOQDEESGLYNNRY--YDPLQGRYITQD---PIGLAGMWLYNY 1243  
 QY 1370 CLGPIINESDFSG-----HLSQAWTIGMGIGLILIT-----IATG---GMAIAAGGI 1416  
 Db 1244 PL-NFIIIRMDPLGLNLYQLLDYVWDDSYGTSIDITSGDLISLGHHGLGVAFKKK 1302  
 QY 1417 AAATIASTTTALAFGALSVTSDITSIVSGALEDASPKA--SSILGWVSMGMGAALAES 1474  
 Db 1303 GEMLSDICIVATACCHAGIGGINAITYSVKSLPTSGVSNVSG-VTVGGVGG----- 1356  
 QY 1475 IKGCTKLATHLG-AFAEDGENALLKSTSESRKIKWGVTRSL 1514  
 Db 1357 -----HPAYIVVDVNDP--ESSTESVGIGAGVDASV 1386  
 RESULT 8  
 AB0539  
 Rhs-family protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strai  
 C;Species: Salmonella enterica subsp. enterica serovar Typhi  
 A;Note: this species has also been called Salmonella typhi  
 C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C;Accession: AB0539  
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A;Reference number: AB0502; MUID:21534947; PMID:11677608  
 A;Accession: AB0539  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-843 <PAR>  
 A;Cross-references: GB:AL513382; PIDN:CAD08754.1; PID:gl6501575; GSPDB:GN00176  
 C;Genetics:  
 A;Gene: STY0324

Query Match 4.0%; Score 353.5; DB 2; Length 843;  
 Best Local Similarity 22.8%; Pred. No. 1.6e-11;







QY 1230 -----NRLVSKTDLDCELYR-----ETMLVNEVRNGEMI 1261  
Db 1019 RERDLTGWMSLSRKEPEETWYGDORLTVTQOTRIQTVYQPSFPLRLRIETENGEOA 1078  
QY 1262 RLLR-----TGETIIAQ-----ORASKVLLTGTDGQQS-----VILTSCKONLS 1300  
Db 1079 KARHRSIAEVLQEDTGTPLPAELAVMLGRLERELROGSVEESQOQLAQCLTAEQMGAQ 1138  
QY 1301 QEA-----YSAGKHKSTANDASI-----LGYN 1323  
Db 1139 LEAGVIPERKHLHCHDQGLPLGLISPGRETALTAEDWGNLLSTSAPQIQQSIRFP 1198  
QY 1324 GERADPVGVTGLONGVRSYDPTLMRHTPDPSLPFG-AGGINPYSYCLGDPINRSDPSG 1382  
Db 1199 GQOYDEESGLYNNRY--YDPLQCRVITQD---PIGLEGGWNLVQVPL-NPIEHDPGLG 1252

RESULT 11  
H91236  
RhsH core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C;Accession: H91236  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strains  
A;Reference number: A99629; PMID:21156231; PMID:11258796  
A;Accession: H91236  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1394  
A;Cross-references: GB:BA000007; PID:BA0364340; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: Ecs4864  
C;Superfamily: rhsF protein

Query March 3.8%; Score 341; DB 2; Length 1394;  
Best Local Similarity 21.8%; Pred. No. 1.9e-10;  
Matches 260; Conservative 156; Mismatches 404; Indels 392; Gaps 63;

QY 418 KETW-RSADNSYRSBITETTFDESIGNPLTKVI-----KDKTKIIS-----PS----- 461  
Db 218 RTQTFHRAAGFSGITGT-DGAGHFLVLTQQAQAEARQQAISGGTEPAFPDT 276  
QY 462 -----THWEYPPAGEVDNCPPEP---YGFT-----RFVK 488  
Db 277 LPGYBYGRDNGIRLSAVMLTHDPEYP-----ENLPAALVRYGWTGELAVVYDRSGK 331  
QY 489 KIIQTPYDSEFK-----DDEKFIQVRYSLIGSQSHVTLKIEERHYSATQLLNSTL 539  
Db 332 QVRSFYDDKYGRVVAHRTGRPE--IRYRD---SQRVTEQNPAGLSVT----- 379  
QY 540 FQNTDKSBLGRLKQTEC--TKGENKTYSVVHKFTYTKQDDTLQQSHSIITHDNFTIH 597  
Db 380 YQVEKDRITITSLNRRVHLHQEGG-LKRVRVK-----EHADGSVT 421  
QY 598 RSQVRSRYTGRFSDPTDKDITVQNSYDKLGRLLITLNSGTPYANTLYDYELANLQDD 657  
Db 422 QSQFDA--VGLRQAQDAAGRTYSPDVVTGLITRIT---TPDGRASAFY----- 468  
QY 658 NRPPFVITITDVGNNQRLNEFDGAGRHVSQCLKSD-----GDGKFYIHTQYDEQGRH 712  
Db 469 NHHSQLTSATGPDGLREIREYDEWGLRIQETAPDGDITRYDNPDSLPACATEDATGSR 528  
QY 713 HTSTVSDYLTNGRQQTDPKXVHLSKMSKYDNQVQANTHWSYGVSEKIVTPPI-TLTATK 771  
Db 529 KTWTSRY---GQLLSFTDCSGYVTRYDHFQGMQTAHREBGLSQVRYAYDSRGOLIAVK 585  
QY 772 QIQSNNSNVQTKCVTVTPSQPIQITLTFDEAGHLQSCHTLTRDGDWRVRKETAICQ- 830  
Db 586 DTQGHETRYE-----YNAAGDLTT--VIAPDG-SRNGTQYDAMGKA 623

QY 831 -CTI-----VOYDNNRVIOITLPDGTIVNRKYAPSTDTLITDIRVNGISLQOQTF- 881  
Db 624 ICTQOGLTRSMYDAAGRVIRLTSENGS-----HTFRYDVLRLIQTGDFGRTQRYH 678  
QY 882 -DGLSRILTOSQDGRWVATYSAGNDQCPSTVITPDGFHYQYQPELDDAVLQVASHI 940  
Db 679 HULTGKLRSEDEGLV-----THWYD-EADRLTHRTVKGET 714  
QY 941 TOQSYNPVTGAL--LKAVAQCSLTPIY--PSGR--KMNENIMKKMSYLM----- 988  
Db 715 AERWOYDE-RGWLITDISHISEGHRVT-VHYGYDEKGRLTGERQTVHHPQTEALLAQHEFR 772  
QY 989 ----TLRGLNGYTDLTGTIKLSRDTGHRVTKDSSIKTILNYDDLNRHIGSQVTDLAT 1045  
Db 773 HAYNAQGLAN--RCIPDLSLPAVENLTYG-----SGWLAGMKL 807  
QY 1046 GMLTITTVTEF--DGLNREIGRKLDCSSGHTLIDIQOSWLKTCQLANRIVKLNGLVLRTEQY 1103  
Db 808 GD--TFLVDFTDRDLRHLKTLRRFGRYELTTATVPAQGLQSQHLNS-----LQYDRDY 857  
QY 1104 SYDSRNLNOYKCDGAECPDKYGHISIVTONFTYDIYGNITACHTTFADGTEDHAFKPA 1163  
Db 858 TWNDNGELIR-----ISSPRQ-----TRSYSDSGRLTGCVHTAA-----NLDIRIP 900  
QY 1164 NPTDPC--QLTEVHHHTPD-----MPDN-----IRLKYDKAGRVINITD----- 1200  
Db 901 YATDPAGNELPD-PELHDPDSLMSKWPNDRIARDAHYLYRYDRHGLTEKTLIPEGVIRT 959  
QY 1201 NHGNTENFTYDGLRQNGQ-----GSVGYDPLNRLVSOXQTDLDCEL----- 1244  
Db 960 DDERTHRYHDSQRLHVTYRTOVEEPLVESRYLYDPLGRVAVXVRVRERDLTGWMSLS 1019  
QY 1245 -----YF-----RETMLVNB-----VRNGEMIRLLR-----TGETIIAQRA----- 1276  
Db 1020 RKPQVTWGDGDRLLTITQNDRTIQTQYQGSFTPLIRVATATGELAKTORSLADALQ 1079  
QY 1277 -----SKVLLTGTDSQOSVILTS----- 1294  
Db 1080 QSGGEDGSGVFPVPLVQVMDRLSEILADRVSESRRLWLASCGLTVAQMSQMDPVVTP 1139  
QY 1295 -----DKQNL-----SOEA-----YSAGKHKSTAND-----ASILGYNGERADPV 1330  
Db 1140 ARKHLVHCHDRGLPLALISKEGATEWCAEYDEWGNLNEENPHQLQOLIRLPQOQYDEE 1199  
QY 1331 SGVTHLNGVRSYDPTLMRHTPDPSLPFG-AGGINPYSYCLGDPINRSDPSGHLWQAW 1369  
Db 1200 SGLYV--NRHRYDPLQGRYITQD---PIGLKGGWNLVYTP-SPVNGMDPLGLYEFKSK 1253  
QY 1390 TGIGMGAGLLL 1401  
Db 1254 NIDDIGIFALAM 1265

RESULT 12  
E65145  
rhsB protein precursor - Escherichia coli (strain K-12)  
C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C;Accession: E65145; S47701; B36902; A30092; I54935  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; PMID:97426617; PMID:9278503  
A;Accession: E65145  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1411  
A;Cross-references: GB:AE000424; GB:U00096; NID:G9367230; PIDN:AACT6507.1; PID:gl789894  
A;Experimental source: strain K-12, substrain MG1655  
R;Plunkett, G.  
submitted to the EMBL Data Library, March 1994

A:Reference number: S47666  
 A:Accession: S47701  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1129, 'Q', 1131-1411 <BLU>  
 A:Cross-references: EMBL:U00039; NID:946582; PIDN:AAB18457.1; PID:9466618  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.  
 J. Bacteriol. 175, 2799-2808, 1993  
 A:Title: Rhs elements of *Escherichia coli* K-12: complex composites of shared and unique  
 A:Reference number: A36902; MUID:93259920; PMID:8387990  
 A:Accession: B36902  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1397-1411 <ZHA>  
 A:Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBI:P.132075)  
 R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
 J. Bacteriol. 171, 636-642, 1989  
 A:Title: Structure of the rhaA locus from *Escherichia coli* K-12 and comparison of rhaA  
 A:Reference number: 154935; MUID:9004253; PMID:2403547  
 A:Accession: 154935  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1221-1411 <RS>  
 A:Cross-references: GB:M29717; NID:9147622; PID:9147623  
 C:Comment: the rha core consist of two distinct parts: a large N-terminal core that is  
 C:Genetics:  
 A:Gene: rnsB  
 A:Map position: 77 min  
 C:Superfamily: rnsB protein  
 C:Keywords: transmembrane protein  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-1411/Product: rnsB protein #status predicted <MAN>  
 F:28-55/Domain: transmembrane #status predicted <TM>

Query Match 3.8%; Score 340.5; DB 2; Length 1411;  
 Best Local Similarity 20.4%; Pred. No. 2.1e-10;  
 Matches 287; Conservative 172; Mismatches 459; Indels 490; Gaps 70;

QY	418	KTEW-RGADNSYRSEITETTFDESIGNPLTKVI-----KDKTKTKIIS-----PS-----	461
DB	218	KTQTHREAAGEFSGEITGVTDGAWHFRVLVLTQARAEARQQAISGTEFSAFDDT	276
QY	462	-----THWYYPAGEVDNCPPEP---YGFT-----	
DB	277	LPFYTEYGRDNGIRLSAVWLTHDPEYP-----ENLPAALVRYGWTGPRGELAVVYDRSGK	331
QY	489	KIIOPPYDFEFK-----DDPEKIQYRSLIGSQSHVTLKIBERHYSATQLLNSTL	539
DB	332	QVRSFYDDKYGRWVAHRTGPE--IRYRD---SDGRVTBQLNAPAGLSYI-----	379
QY	540	FOYNTDKSELGRLLKQTEC--TKGENKTVSVVHKFTYTKQDPTLOQSHSIITHNPTTH	597
DB	380	YQYKDRITITDLSLRREVLTQGEAG-LKRVVKK-----EHADGSVT	421
QY	598	RSQVRSYTGSLFSDTKDKIVTQMSYDKLGRLLTRTLNSGTVPANTLYDYELNNLQDD	657
DB	422	QSQFPA--VGRLRATDAAGRTTEYSDDVTGLITRIT---TPDGRASAFYNNHNQ---	473
QY	658	NRPFVITTTDVGNGQRLNEPDGAGRVSCCLKDSD-----GQKPYTHITQYDEQGRH	712
DB	474	-----LTSATGPDGLELRREYDELGRLIQETAPDGDITRYRDNPHSDLPFCATEDATGR	528
QY	713	HTSYSDYLITNGRQOTPDVKVHLSMSKSYDNWGOIANTHWSYGVSEKITVDPT--TLTATK	771
DB	529	KTMVWSRY--GQLLSFTDCSGYTRYDHRFGQMTAVHREGLSQRAYDSRGQLIAVK	585

QY	772	QLQSNNNVQTGKEVTTTTPSQOPIQITLDFEAGHLQ-----SCHTLTRCDWDRVKE	824
DB	586	DTQGHETRYE-----YNIAGDITAVIAPDGSNRNGTQYDANGKAVT	626
QY	825	TDALCOCTITQYDNNRVIOITLPDGTIVNRKAPFSTDTLITDIRVNGISLQQQTF--D	882
DB	627	TQG-GLTRSMEDYDAAGRVIRLTSENGS-----HTTFRYDVLDRLIQETGFGDGTQRYHHD	680
QY	883	GLSLTQSDQGRVWATYTSAGNDQCPSTVITPDGFIHYQVQPELDDAVLQVANSNEITQ	942
DB	681	LTGKLIRSEDEGLV-----THWHYD-EADRLTHRTVRKGTAE	716
QY	943	QFSYNPVTGAL--LKAVAEQSLTPIYV---PSGRL--KMNINDMKMSYLM-----	988
DB	717	RWOYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHPTQTEALLWQHETRA	774
QY	989	TLRGLNGYDITLGTICKISRDTHGRVTOIKDSSIKTILNYDDLNHRHIGSVQVTDLATOH	1047
DB	775	YNAOGLAN--RCIPDSDLPVAVWILTYG-----SGVLAKMKLGD-----	809
QY	1048	MLTTTVEF--DGLNREIGRKLCDSSGHTLIDQOSWLKTOQLANRIVKLVGLQTEQSY	1105
DB	810	--TPLVEYTRDLHRETLR-----SFGRYELTAYTPAGLOQSO--HLNSLLS-DRDYTW	859
QY	1106	DSRNLNQKCDGACPTDKYGHISVTQNTFYDIYGNITACHTTFADGTEDHATKFNAP	1165
DB	860	NDNGELIR-----ISSRQ-----TRSYSTTGRLTGVHTTAA-----NLDIRIPVA	902
QY	1166	TDPG--QLTEVHETHPD-----MPDN-----IRLYDKAGRVINITD-----NH	1202
DB	903	TDPAGNELPD-PELHPDSTLSMPDNRIARDAHLYVDRHGLRTEKTDLIPEGVIRTD	961
QY	1203	NTNFYDILGRLOQC-----GSVGYDPLNRLVSKQTDPLDCEL-----	1244
DB	962	ERTHYHYDSOHLRVHVTQYEEPLVESRYLYDPGRVAVKRVWRREKDLTGWMSLRK	1021
QY	1245	--YY-----RETMLVNE-----VRNGEMIRLLR-----TGETIIAQORA-----	1276
DB	1022	PQVTWYMGDGRLLTIQNDRSRIQTIYQPGSFPLIRVETATGELAKTORSLADALQOS	1081
QY	1277	-----SKLLTGTDSQSVILTSQKQLSQA-----	1303
DB	1082	GGEDGSGVFPFVLVQMLDESEILLA--DRVSEESRRWLASCGLTVEQMNQMDPVVT	1138
QY	1304	-----XSAKGKHKSTAND--ASILGYNGERADP	1329
DB	1139	PARKIHLVHCDHRGLPLALISTEGATAWCAEYDEWGNLLNEENPHQLQLRLPGQYDE	1198
QY	1330	VSGYTHLNGYRSYDPTLRFHTPDSLSBPG-AGGINPYSCYGLDINSDPSGHLWSQA	1388
DB	1199	ESGLY--NNHRYDPLQGRYITQD--PIGLKGNWLYGYQL-NPISIDIDPLGLSMWED	1252
QY	1389	WTGIMGAGIAGLLTIATGNGWAIAGAAIAAATSTTALAFGALSVTSDITSIVSGALE	1448
DB	1253	-----AKSG--ACTNGLCGTISA-----MIGDPKDFSDIDSTAY	1283
QY	1449	DASPKASSILGWVSMGMGAAGLAESAIGKTKLATHGAFADGENALLKSSSESSIRKW	1508
DB	1284	DALNKINS-----QSICEDKEFA-----	1301
QY	1509	GVTRSLDREIVRNEEGOVINKHSRGYTDNFMGKEQ-----AILVHGDKDGFYHTEG	1561
DB	1302	-----GLICKNSGRYFSTAPNRRGERKSGYSPFNPCPNCTEKVSAYHTHG	1346
QY	1562	NKNGK--GPYTRHTPEQLVYLKDNKI	1587
DB	1347	ADSHGEWDEIFSGKDEKIVK-SKONNI	1373

RESULT 13

H64780

rnsB protein precursor - *Escherichia coli* (strain K-12)

C:Species: Escherichia coli  
 C:Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text\_change 01-Mar-2002  
 C:Accession: H64780; J06025; B00092; I69401; S16026  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: H64780  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1426 <BLAT>  
 A:Cross-references: GB:AB000156; GB:U00096; NID:G1786705; PIDN:AAC73599.1; PID:G1786706;  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Sadosky, A.B.; Gray, J.A.; Hill, C.W.  
 Nucleic Acids Res. 19, 7177-7183, 1991  
 A:Title: The rhd-E subfamily of Escherichia coli K-12.  
 A:Reference number: J06025; MUID:92115567; PMID:1766878  
 A:Accession: J06025  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-378 'A', 380-1166 'G', 1168-1426 <SNAD>  
 A:Cross-references: EMBL:X60999; NID:G42732; PID:G42733  
 A:Experimental source: strain K-12  
 R:Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
 J. Bacteriol. 171, 636-642, 1989  
 A:Title: The rhd-E subfamily of Escherichia coli K-12.  
 A:Reference number: A91501; MUID:89123133; PMID:2644231  
 A:Accession: B30092  
 A:Molecule type: DNA  
 A:Residues: 1-100 <SA2>  
 A:Cross-references: GB:M21764; GB:J04224; NID:G147646; PIDN:AAA24542.1; PID:G147649  
 R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
 J. Bacteriol. 172, 446-456, 1990  
 A:Title: Structure of the rhdA locus from Escherichia coli K-12 and comparison of rhdA  
 A:Reference number: I54935; MUID:90094253; PMID:2403547  
 A:Accession: I69401  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1232-1426 <RES>  
 A:Cross-references: GB:M29719; NID:G147644; PIDN:AAA24541.1; PID:G147645  
 C:Comment: the rhd core consist of two distinct parts: a large N-terminal core that is c  
 C:Genetics:  
 A:Gene: rhd  
 C:Superfamily: rhdF protein  
 C:Keywords: transmembrane protein  
 F:1-26/Domain: signal sequence #status predicted <SIG>  
 F:27-1426/Product: rhd protein #status predicted <MAT>  
 F:28-55/Domain: transmembrane #status predicted <TM>  
 Query Match 3.8%; Score 339; DB 2; Length 1426;  
 Best Local Similarity 20.2%; Pred. No. 2.5e-10;  
 Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;  
 QY 572 KFTYTKQDDFLQOQSHSTTHDFTIHRQVRSRYTGRFLSDTDTKDIVTQMSYDKLGRLL 631  
 DB 317 RYIYTAGELL----AVYDSNTQVAFYDAQHPGMVAHYAGPEMRYDYDTGRVV 372  
 QY 632 TRTNGSTPVANTLYDYELNNLQDNRPPFVITTTD-----EQRHHTSYSDYL 721  
 DB 373 EQLNPAG-----LSYRY-----LYEQDR-----ITVTDLSNRREVLHTEGGAGLKRWKKE 418  
 QY 669 -VNGNQLRNEFDGAGRHVSQCLKSDGDKFTHIQDYD-----EQRHHTSYSDYL 721  
 DB 419 LADGSTRSYDAAGRLTAQ--TDAAGRTEYGLNVSGDITDITPDGRETQFYND-- 474  
 QY 722 TNGRQQT---DPPKVLHSMKSYDNNWGQIANHWSYGVSEKITVDPITLTATKQLQSNRN 778  
 DB 475 --GNQLTAVVSPD--GLSRREYDEPGLV-----SE 502  
 QY 779 NVQTKGKVT--TYTPSQOPIQITLFDGAGLQCHLITRCGDRVRKETAICQCTIYQVD 837  
 DB 503 TSKSGETVRYRYDDAHSELFPATTTDATG---STROMTWSRYGQLLAFTDCSGYQTRYEYD 559

RESULT 14

E85509

hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9

C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E85509  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamianos, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85509  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1404 <STO>  
A:Cross-references: GB:AE005174; NID:gl2512977; PIDN:AAG54537.1; GSPDB:GN00145; UWGP:Z02  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z0268  
C:Superfamily: rhaF protein

Query Match 3.88; Score 336.5; DB 2; Length 1404;  
Best Local Similarity 20.08; Pred. No. 3.4e-10;  
Matches 252; Conservative 130; Mismatches 373; Indels 505; Gaps 58;

Qy 376 YHL-LTSECKQONGYIOTTAYAYAIIGHNFDSPQSQFOLPKT-KTETWRSADNSYRSI 433  
Db 245 FHLVLTQAQAEVFRQRAATSLSSPAGPR--SASSSLVFPDPLPAGTEYGADNGIRLEA 302  
Qy 434 TETTFDES-----GNPLTKVIXDKTKKILSPSTHWEYPPAGEV---DNCPPPYGF 483  
Db 303 VMLTHDPAYPDLPAAPLARYT-----YTASGELRAYVDRSGTQVRGF 345  
Qy 484 TRFKXIIQTPDSE-----FKDDPEKFIQYRSLIGSQSHVTLKIEERHYGATOL 534  
Db 346 A-----YDAEHAGRVVAHYGRPE--RYRYDDTG-----RVTEL 379  
Qy 535 LN-----STLFQYNTDKSELGRLLKQTEC--TKGENKTVSVVHKFTYTKQDDTLQOSH 588  
Db 380 VNPEGLDYREYGDQRTVITDLSNRREVLYTEGEGG-LKRWVK-----422  
Qy 589 TTHDNFTIHSOVSRYTGRFSDTDKDIVTQMSYDKGLRLTRILNSGTPYANTLYD 648  
Db 423 -EHADGSGITSE-----YDEAGRLKAQT-----444  
Qy 649 YELNQLQDNRPPFVITTTDVNGNLNREFDGRHVSQCLKDSG-----DGKFTT 700  
Db 445 -----DAAGRRTEYSLHMASGAVTAVTGDGR--T 472  
Qy 701 IHTQYDEQGRHTSTYSYDLTNGRQOTDPDKVHLSMSKSYDNWQGIANTHWSYGVSEKI 760  
Db 473 VR-YGYNSQRQVTSVTPDGLRSREYDEKRLAAETSRGE-----TTRYSD-----520  
Qy 761 TVDPTITLATKQLOSNNVQTKREVITYTPSQOPIQITLFDGAGHLQSCHTLTRGDWR 820  
Db 521 --DP-----ASELPTGIQDATGSTKQ-----AMSR 544  
Qy 821 ---VRKETAIGCTIYQYDYNRVYIQTLPDGTIVNRKYAPSTDTLDIRVNGISLG 877  
Db 545 YGQLLTFTDCSGYTRYEYDRYGOQIAVHREGLSTYSYNP-----RG 588  
Qy 878 QOTFDGLSRITQSQGRVWAYTYTSAGNDQCPSTVITPDQGFHYQYQPELDDA---VLQ 934  
Db 589 Q-----LVSKDAQGRTRYEYSAAGDL--TAIVAPDGRSEIQY-----DAWGKAVS 634  
Qy 935 VASNEITQCPSPNPVTGALLKVAEGOSLTPPIYPSGRLLKMNINMKMSYLTWLEGLE 994  
Db 635 TTQGLTRSMGYDAAGRITVLTNENGQSSTFRYPDVPDLRTEQRFQDGTORY-----686  
Qy 995 NGYTDLTGTI-----QKISRTHGRVTOIKDSS-- 1022  
Db 687 --HYDLTKGLTQSEDEGLVTLWHYDASDRITHRVNCGDPAEQWYDEHGWLTLTSHSEG 744  
Qy 1023 --IKTTLNYYDLNRIHISQ--VTDLATGHMI-----1049  
Db 745 HRVSVHYGYDDKGLRTGERTVENPETGEMLWEHETGHAYSEQLATRQSPDGLPPVEWL 804

Qy 1050 -----TTTVEF--DGLNREIGRKLDSSGHTLDTQQSWLKTLQOLANRVLK 1093  
Db 805 TYGSGYLAGMKLGCTPLVEYMRDLRHRETARSF--GGEAYELATAWNTSGQURSRLNL 861  
Qy 1094 NGVLQRTQEQSYDSRNLNQKCDGAECPDKYGHSTVQNTFYDYIYGNITACHTTTPADG 1153  
Db 862 P--QLDRDYDNDNGQL--IRISGPQ-----ESREYRYSDDTGRLTGVHTTAANL 906  
Qy 1154 TED--HATFKPAMP-TDPCQLTEVHHTHPD-----MPDNRL-----KYDKAGRVIN 1197  
Db 907 DIDIPYATDPAGNRLPDF-----ELHDDSTLTAWPDN-RIAEADAHVYVRYDEYGRLEA 958  
Qy 1198 ITD-----NHGNTENTYDTLGLL-----QNGQGSV---YGYDPL-----1229  
Db 959 KTDRIPEGVIRMDERTHHYHDSQHLRVFHTRIHQGEPOVESRYLYDPLGRRTGKEVVR 1018  
Qy 1230 -----NRLVSOKTDTLDCELYR-----ETMLVNEVRNGEMI 1261  
Db 1019 RERDLTGWMSLSRKPETEWTYWGWDGRLTTVQTQOTRIQTVYQPGSFTPLLRITENGEQA 1078  
Qy 1262 RLLR-----TGETIIAQ-----QRASKVLLTGTDSQOS-----1289  
Db 1079 KARHSLAEVLQEDTGTVTPAEILAVMLGRLERLRQGSVSESOQWLACGLTAEQMAAQ 1138  
Qy 1290 -----VILTSDKONLSQEAYSAYGK---HKSTANDASILGYN 1323  
Db 1139 LEAEVYPERKHLHYCHDRGLPLALISPEGETAWQGEYDEWGNLLGETSACHLQQSLRLP 1198  
Qy 1324 GERADPVGTVHLNGVRSYDPTLMRPHTPDLSLSPFG-AGINPYSYCLGDPINRSPSG 1382  
Db 1199 CQCYDEESGLYNNRY--YDPLQGRYITQD---PILEGGWNDLYQPL-NPIEHIDPLG 1252

RESULT 15  
C65159  
rhaA protein precursor - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: C65159; S47814; I69402  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Co  
A: Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: C65159  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1377 <BLAT>  
A:Cross-references: GB:AE000437; GB:U00096; NID:92367249; PIDN:AAC76617.1; PID:gi790020;  
R:Feulener, G.; Gray, J.A.; Kirschmann, J.A.; Lerner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
J. Bacteriol. 172, 446-456, 1990  
A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w  
A:Reference number: I54935; MUID:90094253; PMID:2403547  
A:Accession: I69402  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-864, 'M', 866-1377 <RES>  
A:Experimental source: strain K-12, substrain MGL655  
A:Cross-references: GB:M29716; NID:gl47613; PID:gl47614  
A:Experimental source: strain K-12  
C:Comment: the rha core consist of two distinct parts: a large N-terminal core that is co  
C:Genetics:  
A:Gene: rhaA

Search completed: July 3, 2004, 05:47:37  
Job time: 38 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:42:23 ; Search time 19 Seconds  
(without alignments)  
4584.914 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879

Sequence: 1 VYIKFLKFRITMSDNEF.....PRKILGRTEKTVKPTPRP 1673

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	432	4.9	2334	1 WAPA_BACSU	Q07833 bacillus su
2	342	3.9	1397	1 RHSC_ECOLI	P16918 escherichia
3	341.5	3.8	1411	1 RHSE_ECOLI	P16917 escherichia
4	339	3.7	1426	1 RHSD_ECOLI	P16919 escherichia
5	331.5	3.7	1377	1 RHSA_ECOLI	P16916 escherichia
6	220.5	2.5	1183	1 CNA_STAAU	Q53654 staphylococ
7	216.5	2.4	2894	1 YD96_METJA	Q58791 methanococ
8	206	2.3	1902	1 P2P_IACLC	P15293 lactococcus
9	205.5	2.3	1902	1 P3P_IACLC	P15292 lactococcus
10	201	2.3	1902	1 P1P_IACLC	P16271 lactococcus
11	200.5	2.3	1902	1 P2P_LACPA	Q02470 lactobacill
12	198	2.2	2660	1 YEEJ_ECO57	Q04877 escherichia
13	196	2.2	2867	1 RBP2_PLAVB	Q00799 plasmodium
14	188.5	2.1	682	1 RHSE_ECOLI	P24211 escherichia
15	186.5	2.1	2003	1 YHSA_ECOLI	P33666 escherichia
16	185.5	2.1	2358	1 YEEJ_ECOLI	P76347 escherichia
17	185	2.1	1861	1 APU_THETU	P38536 t anylepull
18	183	2.1	3381	1 PGCV_BOVIN	P81282 bos taurus
19	181	2.0	1228	1 SLAP_BACST	P35825 bacillus st
20	180.5	2.0	1577	1 YHSA_PROMI	P16466 proteus mir
21	178.5	2.0	3178	1 YS89_CABEL	Q09624 caenorhabdi
22	174.5	2.0	1803	1 YLJ3_YEAST	P47024 bacillarmyc
23	173.5	2.0	1257	1 CCAA_BACTU	Q45754 bacillus th
24	173	1.9	1829	1 FRPC_NEIMB	Q91975 neisseria m
25	172	1.9	1829	1 FRPC_NEIMC	P55127 neisseria m
26	171.5	1.9	1592	1 GTF2_STRDO	P27470 streptococ
27	170.5	1.9	3097	1 CADN_DROVE	O15943 drosophila
28	169.5	1.9	2314	1 PTP2_HUMAN	P23471 homo sapien
29	168	1.9	1744	1 TANX_XENLA	Q01550 xenopus lae
30	165	1.9	828	1 PMFC_PROMI	P53514 proteus mir
31	163	1.8	1698	1 41_DROME	Q04877 drosophila
32	163	1.8	3358	1 PGCV_MOUSE	Q02059 mus musculu
33	161	1.8	1953	1 BIGA_SALTY	P25927 salmonella

34 160.5 1.8 1018 1 FNBA\_STAAU P14738 staphylococ  
35 160.5 1.8 1560 1 TENN\_MOUSE Q00271 mus musculu  
36 160.5 1.8 2021 1 OMPA\_RICCN Q52657 rickettsia  
37 160 1.8 1569 1 YFUA\_ECOLI P52143 escherichia  
38 159.5 1.8 1523 1 DFOL\_THEFM P74918 thermococcu  
39 159 1.8 1409 1 HAP1\_HAEIN P44596 haemophilus  
40 159 1.8 1723 1 PM20\_CHLPN Q92812 chlamydia p  
41 158.5 1.8 1656 1 OMPB\_RICJA C06653 r outer mem  
42 158 1.8 1812 1 SUIS\_SUNWU O62653 suncus muri  
43 157 1.8 2145 1 CYAA\_PODAN Q01513 podospora a  
44 156.5 1.8 1271 1 Y338\_MYCGE P47580 mycoplasma  
45 156.5 1.8 1609 1 FIG2\_YEAST P25653 saccharomyc

#### ALIGNMENTS

RESULT 1  
WAPA\_BACSU  
ID WAPA\_BACSU STANDARD; PRT; 2334 AA.  
AC Q07833;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Wall-associated protein precursor.  
GN WAPA OR NI7G OR BSU9230.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]\_TaxID=1423;  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=93302506; PubMed=8316082;  
RA "Molecular analysis of three major wall-associated proteins of  
RT Bacillus subtilis 168: evidence for processing of the product of a  
RT gene encoding a 258 kDa precursor two-domain ligand-binding  
RT protein."  
RL Mol. Microbiol. 8:299-310(1993).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=168 / BGSC1A1;  
RX MEDLINE=95319088; PubMed=7704263;  
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.  
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis  
RT genome containing the hut and wapa loci."  
RL Microbiology 141:337-343(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / BGSC1A1;  
RX MEDLINE=97124196; PubMed=8969509;  
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,  
Miwa Y., Fujita Y.  
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome  
RT containing the lic and cel loci, and creation of a 177 kb contig  
RT covering the gnt-sacxy region."  
RL Microbiology 142:3113-3123(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,





Db 1542 QANVTYADSGTGIYNAKPSGTQGWRAAVVPTKTPINKYKINXKIDISILFQKSAITGVWFDDI 1601  
Qy 665 -----TTTVANGNQLRNEFDGAGRHVSCLKSDGDKGYFIHTHQVDEQGRHHT 714  
Db 1602 RLIEGSLTKSTYSDNSNGVYVTEDELGVATS-----TDYDETGK-----KTSETDAKGEKTT 1654  
Qy 715 STV--SDYLTNGRQQTDPKRVHLSMSKSVNDWG-QIANT-----HWSGVSEKI-- 760  
Db 1655 YTTDQADQLNMTLSNGTSILH-----SYDKEGNEVSKTIIRAGADQTYFEFVDMGKLVK 1709  
Qy 761 TVDPITLTATKQKSNNSNNVQT-----GKEVTTYTPSQQPIQIOTLPDEAGHLOSCHTLTRD 816  
Db 1710 TTDPLGNVLASEYDANSNLTKTISPNQNEV-----SLSYD 1744  
Qy 817 GWRVRKETDAICQCTIYQYDYNVRIQITLPDGTIVNRKYAPPSFTDLITDIRVNGISL 876  
Db 1745 GTRVRSKSYNGTEKIFITYDKNGN-----ETSVNKEQN-----TT 1781  
Qy 877 GQOTFQGLSLTQSQDGRWVATYSAGNDQCSTVITPDQPIHYQYQPELDDAVLOVA 936  
Db 1782 KKTFFDNKRLTETLDEGSGCTWTPSDSKLTP-----SWIH-----G 1821  
Qy 937 SNEITQOFSNVPTGALLKAVASQSLTPPIYPSGRLKMNINDMKMSVYLTWLRLENG 996  
Db 1822 DQKGTNQFTYN-----KLDQMTIEMKDSSTSSYFDYDEN- 1854  
Qy 997 YTDLTGTIOKISRDTGRVTOIKDSSIKTTLAYDNLNRHIGSOVTLATGHMLTTTVEPD 1056  
Db 1855 -----GNVQ-----TFITGNGGTSFSFDERNLVSSLHIGDKNGDILTESYEV- 1898  
Qy 1057 GLANREIGRKLCDSGHGLTDIQOEWLKTQOLANRIVKLVNGVLQRTQYSDVSRNLNOYKC 1116  
Db 1899 -----DANGNRTTINS-----ASGV-----QVEYKLNQLVK-- 1927  
Qy 1117 DGAECPDKYGHVITQNPYDLYGNTACHTTFADGTE--HATKFNPTDPCOLTEV 1174  
Db 1928 -----ETHDGTVI-----EYTDGFGNKKTV-TTIKGSSTVNASFNIM--QJTKV 1973  
Qy 1175 H-----HTHP-DMPDNI-----RLKYDKAGRVNIINDNHG 1203  
Db 1974 NDESISYDKNGNRSTDGKFTYTWDAEDNLAVTKGDKGDFATYKYDEKGNRIQKTVN-G 2032  
Qy 1204 NTFNTYDTLGRLONGGSGVGDPLNRLVSQKTDILDCELY-YRETMVLVNEVRNGEMTR 1262  
Db 2033 KVTNYFYDG-----DSLNLVYETDADNNVTKSYTYGD-----SQGLLS 2070  
Qy 1263 LLRTGETIIAQARASKVLLLTGTDSSQSVILTSKONLSQEAYSAYGK-HKSTANDA---S 1318  
Db 2071 YTENGKKYFHYNAAGDIIAISDSGTGV-----AKQYDANGNPKTEASDEVKON 2122  
Qy 1319 ILGYNGERADPVSGVTHLNGYRSYDPTLMRFHT--PDSLSPPGAGGINPYSYCLGDPIN 1376  
Db 2123 RYRYAGYQYDEETGLYYLMARY--YEPRNGVFLSLDPDPSGDSGLDQNGYAYGNPNVM 2180  
Qy 1377 RSDPSGHLSCQAWTGICMGTAGLLTTATGMAIAAAG---GIAAIASTSTTALAFA 1432  
Db 2181 NVDPDGH--W-VMLVNVNAGFA-----AYDGYKAYKSGKWKGAANAASNFQPKIFKG 2231  
Qy 1433 LSVTSDITISVSGALEDASPKASSILGWVSMGM-----GAAGLAESAIKGCTKLATHLGF 1488  
Db 2232 ASRAYKFTK-----KAVKITGHTRHGLNQSIGRNG-----GRGVNLRKALNA- 2273  
Qy 1489 AEDGENALLKSTSSRIKGVTRSLDRE--IVNEEGOVTKDHSRGYTDNFMWKGEOAI 1546  
Db 2274 -----VRSPKKVIRQPNKATYGVGKATVNLNKGKVTITAYG-----SS 2312  
Qy 1547 LVHGDKDGLFVHTEGK 1563  
Db 2313 RAKGSKHVFTHTGKGNK 2329

RESULT 2

RHSC\_ECOLI

ID AC RHSC\_ECOLI STANDARD; PRT; 1397 AA.  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE RHSC protein precursor.  
GN RHSC OR B0700.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
RN NCBI\_TaxID=562;  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=93259920; PubMed=8387990;  
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
RT "Rhs elements of Escherichia coli K-12: complex composites of shared  
RL and unique components that have different evolutionary histories.";  
RN J. Bacteriol. 175:2799-2808(1993).  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
RL corresponding to the 12.7-28.0 min region on the linkage map.";  
RN DNA Res. 3:137-155(1996).  
RP SEQUENCE OF 1-100 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=89123133; PubMed=2644231;  
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
RT "Rhs gene family of Escherichia coli K-12.";  
RL J. Bacteriol. 171:636-642(1989).  
RN SEQUENCE OF 1221-1397 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=90094253; PubMed=2403547;  
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
RT "Structure of the rhaA locus from Escherichia coli K-12 and  
RL comparison of rhaA with other members of the rhs multigene family.";  
RN J. Bacteriol. 172:446-456(1990).  
RP REVIEW.  
RC MEDLINE=95020608; PubMed=7934896;  
RA Hill C.W., Sandt C.H., Vlazny D.A.;  
RT "Rhs elements of Escherichia coli: a family of genetic composites  
each encoding a large mosaic protein.";  
RL Mol. Microbiol. 12:865-871(1994).  
CC -!- FUNCTION: Rhs elements have a nonessential function. They may play  
CC an important role in the natural ecology of the cell.  
CC -!- DOMAIN: Each rhs appears to consist of a highly conserved 141 kDa  
CC amino fragment followed by a highly divergent carboxy terminus.  
CC -!- SIMILARITY: BELONGS TO THE RHS FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE RhsB protein precursor.  
GN RhsB OR B3482.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_taxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=93259920; PubMed=8387990;  
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;  
RT "Rhs elements of Escherichia coli K-12: complex composites of shared  
RT and unique components that have different evolutionary histories";  
RL J. Bacteriol. 175:2799-2808(1993).  
RN [2]  
RP REVISION TO 405.  
RA Hill C.W.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=94318500; PubMed=8041620;  
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
RT region from 76.0 to 81.5 minutes";  
RN Nucleic Acids Res. 22:2576-2586(1994).  
EN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [5]  
RP SEQUENCE OF 1-100 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=89123133; PubMed=2644231;  
RA Sadosky A.B., Davidson A., Lin R.J., Hill C.W.;  
RT "rhs gene family of Escherichia coli K-12";  
RL J. Bacteriol. 171:636-642(1989).  
RN [6]  
RP SEQUENCE OF 1221-1411 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=90094253; PubMed=2403547;  
RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.P., Sadosky A.B.,  
RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.;  
RT "Structure of the rhsA locus from Escherichia coli K-12 and  
RT comparison of rhsA with other members of the rhs multigene family";  
RL J. Bacteriol. 172:446-456(1990).  
RN [7]  
RP REVIEW  
RX MEDLINE=95020608; PubMed=7934896;  
RA Hill C.W., Sandt C.H., Vlazny D.A.;  
RT "Rhs elements of Escherichia coli: a family of genetic composites  
RT each encoding a large mosaic protein";  
RL Mol. Microbiol. 12:865-871(1994).  
CC -!- FUNCTION: Rhs elements have a nonessential function. They may play  
CC an important role in the natural ecology of the cell.  
CC -!- DOMAIN: each rhs appears to consist of a highly conserved 141 kDa  
CC amino fragment followed by a highly divergent carboxy terminus.  
CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; L02370; AAC61883.1; -  
DR EMBL; U00039; AAB18457.1; -  
DR EMBL; AE000424; AAC76507.1; -  
DR PIR; E65145; E65145.  
DR EcoGene; EG10847; rhsB.  
DR InterPro; IPR001826; rhsB.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF03527; Rhs; 1.  
DR Pfam; PF05593; Rhs\_repeat; 13.  
DR PRINTS; PRO0394; RhsPROTEIN.  
DR TIGRFAMs; TIGR01643; YD repeat 2x; 15.  
KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 1411 RhsB PROTEIN.  
FT TRANSMEM 28 55 POTENTIAL.  
FT DOMAIN 330 1186 28 X APPROXIMATE TANDEM REPEATS.  
FT REPEAT 330 352 1.  
FT REPEAT 353 374 2.  
FT REPEAT 375 417 3.  
FT REPEAT 418 438 4.  
FT REPEAT 439 460 5.  
FT REPEAT 461 481 6.  
FT REPEAT 482 502 7.  
FT REPEAT 503 525 8.  
FT REPEAT 526 546 9.  
FT REPEAT 547 567 10.  
FT REPEAT 568 588 11.  
FT REPEAT 589 609 12.  
FT REPEAT 610 629 13.  
FT REPEAT 630 650 14.  
FT REPEAT 651 671 15.  
FT REPEAT 672 691 16.  
FT REPEAT 692 711 17.  
FT REPEAT 712 734 18.  
FT REPEAT 735 758 19.  
FT REPEAT 808 828 20.  
FT REPEAT 829 850 21.  
FT REPEAT 851 871 22.  
FT REPEAT 872 894 23.  
FT REPEAT 895 930 24.  
FT REPEAT 931 959 25.  
FT REPEAT 960 984 26.  
FT REPEAT 985 1019 27.  
FT REPEAT 1162 1186 28.  
FT CONFLICT 1130 1130 Q -> K (IN REF. 4).  
SQ SEQUENCE 1411 AA; 159394 MW; 567AC4EE713D9E07 CRC64;  
  
Query Match 3.8%; Score 341.5; DB 1; Length 1411;  
Best Local Similarity 20.4%; Pred. No. 1.3e-10; Indels 484; Gaps 70;  
Matches 287; Conservative 172; Mismatches 462;  
  
QY 418 KETW-RSADNSYRSEITETTFESGNPLTKVI-----KDKTKIIS-----PS----- 461  
DB 218 RTCTFHRAAGFSGEITGVT-DGAWRHFLVLTTCQAQAEAEARCAISGGTSPSAPDPT 276  
QY 462 -----TWYYPAGEVDNCPPEP---YGET-----RFLK 488  
DB 277 LFGYTEYGRDNGIRLSAVWLTHDPEYP-----ENLPAELVRYGWTFRGELAVVYDRSGK 331  
QY 489 KIIQTYDSEFK-----DDPEKFIQYRSLIGSQSHVTKIEERHYSATQLINSL 539  
DB 332 QVRSFTYDDKYGRVVAHRTGEP--IFYRD---SDGRVTEQLNPAGLSVT----- 379  
QY 540 FQYNTDKSELGRLKQTEC--TKGNGKTYSVVHKFTYTKQDRTLQOQSHSIHNFTH 597  
DB 380 YQVEKDRITITSLDRREVLHTQGEAG-LKRVVK-----EHADGSVT 421  
QY 598 RSQVRSRYGRFLSDDTKDVIYQSYDKLGRLLTTLNSGTPYANTLYDYELANLQDD 657  
DB 422 QSQFDA--VGRRAQTDAAGRTESPDVVTGLITRIT---TPDGEASAFYHNHNO---- 473  
QY 658 NRPPFVITTTDVGNGQLRNEFDGAGHVSQCLKSD-----GDGKPYTHITQYDEQGRH 712



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X60999; CAA43314.1; --  
 DR EMBL; AE000156; AAC73593.1; --  
 DR EMBL; U82664; AAB40251.1; --  
 DR EMBL; AF044500; AAC32467.1; --  
 DR EMBL; M21764; AAA24542.1; --  
 DR EMBL; M29719; AAA24541.1; --  
 DR PIR; H64780; H64780.  
 DR EcoGene; EG10849; rhsD.  
 DR InterPro; IPR001826; RHS.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF03527; RHS; 1.  
 DR Pfam; PF05593; RHS; repeat; 11.  
 DR PRINTS; PR00394; RHS; PROTEIN.  
 DR TIGRFAMs; TIGR01643; YD repeat 2x; 16.  
 KW Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1426  
 FT TRANSMEM 28 55  
 FT CONFLICT 379 379  
 FT CONFLICT 1167 1167  
 FT SEQUENCE 1426 AA; 159724 MW; 122D697AA449BC2F CRC64;

Query Match 3.8%; Score 339; DB 1; Length 1426;  
 Best Local Similarity 20.2%; Pred. No. 1.8e-10;  
 Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;  
 572 KFTYTKDDDTQQSHSITTHDNFIHRSQVRSYRGLFSDTDKDIVTQMSYDKLGRLL 631  
 317 RYVTEAGELL---AVYDRSNTQVRAFTYDAQHPGWAHVRAGREMYRVDIGRVV 372  
 632 TRTNSGTPYANTITYYELNLODDNRPPVITTD----- 668  
 373 EQLPAG-----LSYRY---LYEQDR---IVTDSLNRREVLHTEGGAGLKRVRVKE 418  
 669 -VNGQLRNEFDGAGRHVSQCLKSDGDKFYTHITQYD-----EQGRHHSTYSYDL 721  
 419 LADSVTRSGVDAAGRTAQ--TDAAGRTEYGLNVVSGDITDITPDGRETIFYND-- 474  
 722 TNGRQOT---DPPDKVHLSMSKSYDNWCQIANHWSYGVSEKITVDPITLTKQLQSN 778  
 475 --GNQLTAVVSPD--GLESRREYDEPGRV-----SE 502  
 779 NVOTGKEVT-TYTPSQPIQITLFDAGHLQSCHTLTRDGDWRVRKETDAIGQCTIYQYD 837  
 503 TSRSGETVRVRYDDAHSELPAITTDATG---STROMTWSRYGQLLAFTDCSGYQTRYED 559  
 838 NYARVIQITLFDGTIVNRKYAPSTDLITDIRVNGISLQQQTFDGLSRLTSQDQ-GRV 896  
 560 RFGQM-----TAVHRE-----EGISL-YRYDNRGLTSVKDAQRE 595  
 897 WAYTYSACNQCSPTVTPGQTHIYQYQ----- 925  
 596 TRYENAAAGL--TAVITPDGNRSETQYDAGWKAIVSTQGLRSMEDYDAAGRVISLTNE 653  
 926 -----PELDDAVLQVANSNETTQOFSYNPVTGALLKAVARGQSITPIYYPGSLKME 976  
 654 NGSHVSFSDALDELVOQGFGDGTQRYHYD-ITGKLTQSEDEGLVILWYDESDRITH 712  
 977 NINDMKQSYLWTLRGLNGYTDLTGFIQKISRDTHGRVTQIKDSS-----IKTLLNYDDL 1032  
 713 TVNGEPAEQWY-----DGHGWLTDIHLSEGHVAVHYGYDDK 751  
 1033 NRHIG--SQVTDLATGHML----- 1049  
 752 GRLTGECQTVENPETGELLWQHETKAYNEQGLANRVTPDSLPPVWLTGYSGYLGMKL 911

QY 1050 ---TTTVEF--DGLNREIGRKLCDSSGH--TLDIQSQWLKTLQQLANR----- 1089  
 DB 812 GGTFLVEYTRDLRHRETIVRSFGSMAGSNAAYELTSTYTPAGQLSQHLSLVYDRDYGWS 871  
 QY 1090 ---IVKLANGVLQRTQYSYDSNRNLNQVKCDGAC-----PTDKYVGHSI----- 1130  
 DB 872 DNGDLVRISGPRO-TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGRLDPDELHPDS 930  
 QY 1131 -----VTQN-----FTYDIYGNITACHTTTFADG---TEDHATKFKFANTDPDQLTEV 1174  
 DB 931 TLTWPDNRIADAHYVVRHDEYGRLTXTKTRIPAGVIRTDDBERT----- 975  
 QY 1175 HHTHPDMPDNIRLKYDXKGRVINITD-NHGN---TENFTYDILGR----- 1215  
 DB 976 HHYH-----YDSQHLRVFYTRIQEGLPVLVSRYDLPGLRRMAKRVRRERDLTG 1025  
 QY 1216 ---LQNGQSVYGYDPLNRLYSQKTDITLDCELYR----- 1247  
 DB 1026 WNSLSRKPEVTWYMGD-DRLTTVQITDTRITQTVTEPGSFTPLIRVETENGEREKAQRS 1084  
 QY 1248 ---ETMLVNEVRNG-----EMIRLL-RGETIIL-----QORASKV 1279  
 DB 1085 LAETIQEGSENGHGVVFPFAPLVRLLDRLEESIRADRVSSSESRAMLACGLTVEQLARQV 1144  
 QY 1280 LLTGTDSQS-----VILTSDKONLSQEA-YSAYGKHKSTANDASILGYN--- 1323  
 DB 1145 EPEYTPARKAHLVCHDRGLPLALISEDGNTAWSAEYDEWGNLNEEPHV--YQPYL 1202  
 QY 1324 -GERADPVSGVTHLNGVRSYDPTLMRPHTPDLSPPFG-AGGINPYSCYCLGDPINRSDPS 1381  
 DB 1203 PQQHDDEESGLY--NRHRYDPLQGRYITQD---PMGLKGGWNLVQYPL-NPLQQIDSM 1256  
 QY 1382 GHLSCQAWTIGMGITAGLLTATGMAIAAAGGIAAATASTTALAFGALSVTSDITS 1441  
 DB 1257 GLL--QTWDDARSG-----ACTGV-----CGVLSRIIGPSKFDSTADAALD----- 1296  
 QY 1442 IVSGALEDASPKASSILGWVSMGAAGLAESAIKXGTKLATHLGAFAEDGENALLKSTS 1501  
 DB 1297 ---ALKETQNRS-----LCNDMEYSGIVCKDTNG-----KYFASKAETDNL--K 1336  
 QY 1502 ESSRIKQWVTRSLDRIVRNEBGOVVKOHSRQYTDNEMCKGFOAILVHGDK--DGFVHT 1559  
 DB 1337 ESYPLURKCPCTGTRVAAYHTHG---ADSHGVDYDEFFSSSDKNLVRSKNNLEAFYLAT 1393  
 QY 1560 EGNKH---NGKGPY 1570  
 DB 1394 PDGRFEALNNKGEY 1407

## RESULT 5

RHSA\_ECOLI  
 ID RHSA\_ECOLI STANDARD; PRT; 1377 AA.  
 AC P16916;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE RhsA protein precursor.  
 GN RhsA OR B3593.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID:562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90094253; PubMed=2403547;  
 RA Feulner G., Gray J.A., Kirschman J.A., Lehner A.F., Sadosky A.B.,  
 RA Vlazny D.A., Zhang J., Zhao S., Hill C.W.,  
 RT "Structure of the rhaA locus from Escherichia coli K-12 and  
 RL comparison of rhaA with other members of the rha multigene family."  
 RN [2]

SEQUENCE FROM N.A.  
 RP STRAIN=K12 / MG1555;  
 RA MEDLINE=94316500; PubMed=8041620;  
 RX Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the *Escherichia coli* genome. V. DNA sequence of the  
 RL region from 76.0 to 81.5 minutes.";  
 RN Nucleic Acids Res. 22:2576-2586(1994).  
 [3]  
 RP REVIEW  
 RX MEDLINE=95020608; PubMed=7934896;  
 RA Hill C.W., Sandt C.H., Vlazny D.A.;  
 RT "Rhs elements of *Escherichia coli*: a family of genetic composites  
 RL each encoding a large mosaic protein.";  
 CC Mol. Microbiol. 12:865-871(1994).  
 CC -!- FUNCTION: Rhs elements have a nonessential function. They may play  
 CC an important role in the natural ecology of the cell.  
 CC -!- DOMAIN: Each rhs appears to consist of a highly conserved 141 kDa  
 CC amino fragment followed by a highly divergent carboxy terminus.  
 CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L19044; AAC95065.1; -;  
 DR EMBL; U00039; AAB18570.1; -;  
 DR EMBL; AE000437; AAC76617.1; -;  
 DR PIR; C65159; C65159.  
 DR EcoGene; EG10846; rhsA.  
 DR InterPro; IPR001826; Rhs.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF03527; Rhs; 1.  
 DR PRINTS; PF00394; RhsRepeat; 12.  
 DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 15.  
 DR Signal; Multigene family; Transmembrane; Repeat; Complete proteome.  
 KW SIGNAL  
 FT CHAIN 1 26  
 FT TRANSMEM 27 1377  
 FT DOMAIN 28 55  
 FT REPEAT 330 1186  
 FT REPEAT 330 352  
 FT REPEAT 353 374  
 FT REPEAT 375 417  
 FT REPEAT 418 438  
 FT REPEAT 439 460  
 FT REPEAT 461 481  
 FT REPEAT 482 502  
 FT REPEAT 503 525  
 FT REPEAT 526 546  
 FT REPEAT 547 567  
 FT REPEAT 568 588  
 FT REPEAT 589 609  
 FT REPEAT 610 629  
 FT REPEAT 630 650  
 FT REPEAT 651 671  
 FT REPEAT 672 691  
 FT REPEAT 692 711  
 FT REPEAT 712 734  
 FT REPEAT 735 758  
 FT REPEAT 759 782  
 FT REPEAT 783 808  
 FT REPEAT 809 828  
 FT REPEAT 829 850  
 FT REPEAT 851 871  
 FT REPEAT 872 894  
 FT REPEAT 895 930  
 FT REPEAT 931 959  
 FT REPEAT 960 984  
 FT REPEAT 985 1019  
 FT REPEAT 1162 1186  
 FT REPEAT 1377 AA; 156320 MW; 21ACA989E74200FE CRC64;  
 SQ SEQUENCE

Query Match 3.7%; Score 331.5; DB 1; Length 1377;  
 Best Local Similarity 20.8%; Pred. No. 4.3e-10;  
 Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;  
 QY 211 LPPGQKEGY-RTELRLNRLNSHNFLSGNENPLTWSFGYTPIGKNGILGQWITSMTAP 269  
 DB 126 LPPGE-DGYSRSESLVLR--GGVAKLDEGHRLAALMQ----- 160  
 QY 270 GGLKETVNY-----NNNQ-----HHFQSANL-----PVLPIVTLMKQVPGAGQ- 310  
 DB 161 ALPEELRLSPHYLATNSPQGPWMLLGCERVPDEVLPAFLPPIVRLTGLVDRFGRT 219  
 QY 311 PAIQAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNGSTEGRRYKDKEGHQIVRIE 370  
 DB 220 QTFHREAAEFGSGEITGVTGAWRH-----FRLVL-----TTOQR----- 255  
 QY 371 RTNNYHLLTSECKQONGYIOTTETAYYAIIGHNFDSPQSFQPKT---KTETWRSADN 427  
 DB 256 -----AEEARQQ-----AISG-----GTEPSAF--PDTLPGYTEYGR--DN 287  
 QY 428 SVRSITETTTDESNGNPLTKVKKTKQKIISPSTHWEYPPAGEVDMCPEP---YGF 484  
 DB 288 GIRLSAVMLTHDPE-----Y-----ENLPAAPLVRGWT 317  
 QY 485 -----RFVKKIIQTPYDSBFK-----DDPEKFIQYRYSLIGSQSHVTKIE 525  
 DB 318 PRGELAVVYDRSGKQVRSFTYDDKYGRWVAHRHTGRPE--IRYRD---SDGRVTEQLN 372  
 QY 526 ERHYSATQLLNSTLFOYNTDKSELGRLLKQTEC--TKGNGKTYSVVHKFYTYTKDDTLQ 583  
 DB 373 PAGLSYT-----YQYKDRITITDSLRREVLHTQGEAG-LKRWVK----- 413  
 QY 584 QSHSITTHDNFTIHRQSVRSRYTGRFSDTDKQIVTQMSYDKLGRLLTRTLNSGTPYAN 643  
 DB 414 -----ERADGSVTQSQDA--VGLRACTDAAGTTEYSVDVWVTLTRIT--TPDGR 462  
 QY 644 TLTYDYELNLODDNRPFFVITTTDNGNQLRNEFDGAGRHSVQCLKQSD-----GDGKF 698  
 DB 463 ASAFYNNHNQ-----LTSATQDGLERREYDELGRLIQETAPDGTIRVRYDNPH 514  
 QY 699 YTHIQQDEQCRHHTSYSDYLTNGROTDKDVHLSMSKSYDNWQGIANTHWSYGVSE 758  
 DB 515 SLDPCATEDATGRKMTWSRY---GQLLSFTDCSGYTRVDHDFGQMTAVHREGLSQ 571  
 QY 759 KITVDPI-TLTATKQLQNSNNVQTKVTTTPSQPIQITLDFEAGHLQ-----SC 810  
 DB 572 YRAVDSRGQLIAVKDTQGHETRYE-----YNIAGDLTAVIAPDGSR 612  
 QY 811 HLTTRDGDWRVRKETDAIGCQCTIYOYDNVNRVLIQITLPGTIIVNRKYAPFSTDTLITDIR 870  
 DB 613 NGTQYDAMGKAVRTTQG-GLTRSMYDAAGRVIRLITSENGS-----HTTFYDVLDRLIQ 666  
 QY 871 VNGISILGQQTF--DGLSRITQSDGGRVWAYTYSAGNDQCPSTVITPQQFIHYOYQBEL 928  
 DB 667 ETGFDGRTORYHHDLTGKLIRSEDEGLV-----THMYD-EA 702  
 QY 929 DRAVLQVAGNEITQOFSYNPVTGAL--LKAVAEGSLTPIYV---PSGRL--KMNINDM 981  
 DB 703 DRLTHRTVKGETAERWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKRLTGERTVHP 760  
 QY 982 KXMSYLW-----TLRGLNGYDLTGTIKISRDTHGRVTQIKDSSIKTLLNYDDLIN 1033  
 DB 761 QTEALLWQHETHRAYNAOGLAN--RCIPDLPVAVWLTVG-----SGYLAGMKLGD-- 809  
 QY 1034 RHIGSQVTLATGCHMLTTTVEP--DGLNREIGRKLDCSSGHSLDLCQSWLTKQQLANRIV 1091  
 DB 810 -----TPLVEYTRDLRHLRETLR-----SFGRYELTATYTPAGQLQSQ-- 846  
 QY 1092 KLVGLQRTQOYSYDSRNLNOVKDGAECPTDKYGHISVITQNTFYDIYGNITACHTTFA 1151  
 DB 847 HLNSLLS--DRDYTWNDNGELIR-----ISSPRQ-----TRSYSTTGRLTGVHTTAA 893

QY 1152 DGTEDHATKFNPTDPC--OLTEVHTHPD-----MPDN-----IRLKYDKAGRVIN 1197  
DB 894 -----NLDIRIPYATDPAGNLPD--PELHPDSTLSMPDNRIARDAHYLYRYDEHGRLE 947  
QY 1198 LTD-----NHGNTENFYDILGRLOQC-----GSVGYDPLNRLVSKTDT 1239  
DB 948 KTDLPEGVITDDBTHRYHDSQRLVHVTRTQYEPLVESRYLDPLGRVAKVR 1007  
QY 1240 LDCEL-----YY-----RETMLVNE-----VRNGEMIRLLR-----TGETI 1270  
DB 1008 RERDLTGWMSLRKPOVTWYGDGDRLLTIQNDRTRIQTIVQPSFTPLRVERATGELA 1067  
QY 1271 IAQRA-----SKULLTDTSCQSVIL-----TS 1294  
DB 1068 KTORSLADALQSGGEGDGSVFPFVLVQMLDRLESEILADRVSESRRLWASCGLTVE 1127  
QY 1295 DKQN-----LSQEA-----YSAYCKHKSTAND---AS 1318  
DB 1128 QMQLQMDPVYTPARKIHLXCHDRGLPLALISKEGTTWCAYDEWGNLNEENPHOLOQ 1187  
QY 1319 ILVNGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSFG-AGGINPYSYCLGDPINR 1377  
DB 1188 LIRLPGQYDRESGLYY--NHRHYVDPLQGRYITQD---PIGLKGGWNFQYPL-NDVTN 1241  
QY 1378 SDPSG 1382  
DB 1242 TDPLG 1246

RESULT 6  
CNA\_STA00 STANDARD; PRT; 1183 AA.  
AC Q53654;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Collagen adhesin precursor.  
CN CNA.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FDA 574;  
RX MEDLINE=92165839; PubMed=1311320;  
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
Lindberg M., Hoesek M.;  
RT "Molecular characterization and expression of a gene encoding a  
Staphylococcus aureus collagen adhesin."  
RL J. Biol. Chem. 267:4766-4772(1992).  
RN [2]  
RP ERRATUM.  
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
Lindberg M., Hoesek M.;  
RL J. Biol. Chem. 269:11672-11672(1994).  
RN [3]  
RP COLLAGEN-BINDING DOMAIN.  
RC STRAIN=FDA 574;  
RX MEDLINE=94032261; PubMed=8218209;  
RA Patti J.M., Boles J.O., Hoesek M.;  
RT "Identification and biochemical characterization of the ligand  
binding domain of the collagen adhesin from Staphylococcus aureus."  
RL Biochemistry 32:11428-11435(1993).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.  
RX MEDLINE=97475225; PubMed=9334749;  
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,  
Moore D., Jin L., Schneider A., Delucas L.J., Hoesek M.,  
Narayana S.V.L.;  
RT "Structure of the collagen-binding domain from a Staphylococcus  
aureus adhesin."  
RL Nat. Struct. Biol. 4:833-838(1997).  
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO

CC COLLAGEN-CONTAINING SUBSTRATA.  
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M81736; AAA20874.1; -  
CC PDB; 1AMX; 24-JUN-98.  
CC PDB; 1D20; 27-SEP-00.  
CC PDB; 1D2P; 27-SEP-00.  
CC InterPro; IPR008966; Adhes\_bact.  
CC InterPro; IPR008454; Cna\_B\_unit.  
CC InterPro; IPR008970; Cna\_B\_unit.  
CC InterPro; IPR008456; Collagen\_bind.  
CC InterPro; IPR001899; Gram\_pos\_anchor.  
CC Pfam; PF05738; Cna\_B; 7.  
CC Pfam; PF05737; Collagen\_bind; 1.  
CC TIGRfam; TIGR01167; LPXTG\_anchor; 1.  
CC PROSITE; PS50847; GRAM\_POS\_ANCHORING; Repeat; Signal; 3D-structure.  
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 1154 COLLAGEN ADHESIN.  
FT PROPEP 1155 1183 REMOVED BY SORTASE (POTENTIAL).  
FT DOMAIN 151 318 COLLAGEN-BINDING.  
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).  
FT REPEAT 533 719 B1.  
FT REPEAT 720 906 B2.  
FT REPEAT 907 1093 B3.  
FT SITE 1151 1155 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 1154 1154 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
FT STRAND 174 179  
FT TURN 182 183  
FT TURN 185 186  
FT STRAND 187 194  
FT TURN 196 197  
FT STRAND 201 201  
FT STRAND 205 211  
FT STRAND 215 228  
FT TURN 229 230  
FT STRAND 232 234  
FT HELIX 239 246  
FT TURN 248 249  
FT STRAND 251 255  
FT TURN 256 259  
FT STRAND 260 265  
FT HELIX 267 270  
FT TURN 271 272  
FT STRAND 273 283  
FT TURN 286 287  
FT STRAND 290 299  
FT STRAND 301 301  
FT TURN 302 303  
FT STRAND 307 311  
FT STRAND 314 317  
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

Query Match 2.5%; Score 220.5; DB 1; Length 1183;  
Best Local Similarity 19.1%; Pred. No. 0.00039;  
Matches 206; Conservative 159; Mismatches 405; Indels 311; Gaps 56;  
QY 299 VTLKQVFCAGQ---PAIQAEYSYTSYHNYGGSGNGIWNKLDNLYGLMTNYGTSR 355  
DB 205 ITRKDIQGGQDLSTLNINVTGTHSNYSGQS-----AITDFKAPGSK 251  
QY 356 RYDKESGHQIVRIERTYNNHLLTSECK-----QQNGYIQTTETAYTAIG----- 402

Db	252	ITVDNTKMTIDVTIPOQYSGVNSYNSINFKYKTIWNEQKQFVWNSQ-AWYQEHGKEEVNGK	310
Qy	403	-----HNFDSPQSQFQPKTKTTWRSANSYSEITETTFBDSGNPLTKYVKDKKTQK	456
Db	311	SPNHTVHNANAGIEGTGKELVQLQ-KDRTKAPIANVKFLSKKD-GSVVKD--NQK	366
Qy	457	LIISPTWHEYPPAGEVDCPEPVGFTRFVKKIQTQPDSEPKDDPEKFIQYRSL--I	514
Db	367	BI-----EITDANGANIKALESQ--DYILKEIAPRYTFDKXE-----YPTFMKDT	414
Qy	515	GSQSHVTL-----KIE-----BRHYSATOLLNSTLF-----QYNT-----DKSELGR	551
Db	415	DNQGVFTTIEAKALEKTKDVSAQKVGEGTKVKTPTVFKLYKQDDNQNTTPVDPKABIKK	474
Qy	552	LLKQTECK-----GENGKYSVVHKETYTKQDDTLQOSSH-----ITHDNFTI	596
Db	475	LEDGT--TKVTWSNLPENDKNGKAIKYLKVEVNAQGEDTTPGTYKKEGLVVTNTEKPI	532
Qy	597	HRSQVRSRYTGRLFSDTDTKD--IVTQMSYDKLG--RLLTRLNSGTPYANTLYDYEL	651
Db	533	ETTSISGE---KWDDXNDQDGKPEKVSVNLLANGKVKTLDTVSET-----NWKYEF	583
Qy	652	NNL-----QDNRPFFVIT-----TTVNGNQLRNEFDGAGRHSQCCKSDGDKGYTI	701
Db	584	KDLPYDEGKIEYTVTEHDVKDYTDINGTTITNKYT-PGETSATVTKWDDNN-----	637
Qy	702	HTQQYDEQGRHHT-----STYSDLTNGSQQDPDPKVLHSMKSYDNMGQIANTHWSYGS	757
Db	638	NQDGKPEIKVELYQDGKATGK-----TAILNESNNW-----THTWTGLD	678
Qy	758	E-----KITVDPITLTPATKQLSNNVOTGKVVTT--YTPS-----	792
Db	679	EKAQGOQVKYVE--ELAKVKGYTHVDNDMGNLIVTNKYTPETTSISGEKWDDXNQ	736
Qy	793	-----QOPIQITLFDAGHLQSHCHTLTRDQWDRVKEITDAIQCTIYOYDNNRVIQITLP	848
Db	737	DGKRPEKVSNNLADGEKVKTLDTVSETNWKYFKD-----LPKDEGKIEYIVTE	788
Qy	849	D-----GTIVNRKYAPFSDTLI-----TDIRV-----NGISLQO-	878
Db	789	DVHKDYTTDINGTTITNKYTPGETSATVTKWDDNNQDGKRPTEIKVELYQDGKATGT	848
Qy	879	-----QTFDGLSRLTQSQ-----DGRVWATYSAGNDQCPSTVITPDGQFIH	921
Db	849	AILNESNNWHTWTGLDEKAKGOQVKYTVVEELTKVGYTHVDNDMGNLIVT-----	901
Qy	922	YOYQPELDAVLQVANSNITQFQSYNPVVTGALLKAVAEQSLTPIYPSPGELKMNNDM	981
Db	902	NKYTPETTSISGEKWDDXNDQDGKRP-EKVSNNLANGEKVKTLDTVSETNWKYEPKDL	960
Qy	982	-----KMSYLWTLRGLENGWTDLTGTQIKLSRTHGRVTOIKDSSIKITLYVDDLNHR	1035
Db	961	PKYDEGKIEYTVTEHDVKDYTDINGT-TINKYTPG-----ETSATVTKWDDNNNQ	1013
Qy	1036	IGSQVTDLATGHMLTTFVEF--DGLNREIGRKLCDSSGHILDIQOS--WLKTOQLANRIV	1091
Db	1014	DGKRPEI-----KVELYQDG-----KATGKTALINESNNWHTWTGLDEKA	1055
Qy	1092	KLVGLQRTQEQYSRNRNLQYKCGAECPTDKYGHSLVTONFT-----YDIYGNITAC	1146
Db	1056	KGOQV-----KYTVDELTKVNGYT--THVDNDMGNLIVTNKYTPKPKNPIY-----	1101
Qy	1147	HTTFADGTEDHATFKFANPTDPCQLTEVHHTHPMDNIRLKYKAGRVINITHNGNTE	1206
Db	1102	-----PEKPKDKTP-----PTKPDHNSNKVKPFPDPKPS-----KVDKDDQP---	1145
Qy	1207	N 1207	
Db	1146	N 1146	
RESULT 7			
YD96_METJA			

ID	YD96_METJA	STANDARD;	PRT; 2894 AA.
AC	Q58791;		
DC	15-MAR-2004 (Rel. 43, Created)		
DT	15-MAR-2004 (Rel. 43, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Hypothetical protein MJ1396.		
GN	MJ1396.		
OS	Methanococcus jannaschii.		
OC	Archaea; Euryarchaeota; Methanococci; Methanococcales;		
OC	Methanocaldococcaceae; Methanocaldococcus.		
OX	NCBI_TaxID=2190;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=JAL-1 / DSM 2661 / ATCC 43067;		
RX	MEDLINE=96337999; PubMed=8688087;		
RA	Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,		
RA	Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,		
RA	Kerlavage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,		
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,		
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Sadow P.W., Nguyen D.,		
RA	Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,		
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,		
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;		
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus		
RT	jannaschii.";		
RL	Science 273:1058-1073(1996).		
CC	!- SIMILARITY: Contains 20 PBI repeats.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U67579; AAB99406.1; -		
DR	PIR; C64474; C64474.		
DR	TIGR; MJ1396;		
DR	InterPro; IPR008985; ConA like lec_gl.		
DR	InterPro; IPR001220; Lectin_legB.		
DR	InterPro; IPR006626; PBI.		
DR	Pfam; PF00139; lectin legB; 1.		
DR	SMART; SM00710; PBI; 20.		
KW	Hydrochemical protein; Transmembrane; Repeat; Complete proteome.		
FT	TRANSMEM 8 28 POTENTIAL.		
FT	REPEAT 543 567 PBI 1.		
FT	REPEAT 2085 2107 PBI 2.		
FT	REPEAT 2135 2156 PBI 3.		
FT	REPEAT 2158 2180 PBI 4.		
FT	REPEAT 2201 2223 PBI 5.		
FT	REPEAT 2224 2244 PBI 6.		
FT	REPEAT 2245 2266 PBI 7.		
FT	REPEAT 2267 2289 PBI 8.		
FT	REPEAT 2290 2311 PBI 9.		
FT	REPEAT 2341 2363 PBI 10.		
FT	REPEAT 2367 2389 PBI 11.		
FT	REPEAT 2390 2419 PBI 12.		
FT	REPEAT 2422 2444 PBI 13.		
FT	REPEAT 2455 2477 PBI 14.		
FT	REPEAT 2479 2501 PBI 15.		
FT	REPEAT 2512 2542 PBI 16.		
FT	REPEAT 2550 2582 PBI 17.		
FT	REPEAT 2589 2611 PBI 18.		
FT	REPEAT 2612 2633 PBI 19.		
FT	REPEAT 2638 2660 PBI 20.		
SQ	SEQUENCE 2894 AA; 322361 MW; 8B64F786537A16DF CRC64;		
Query Match 2.4%; Score 216.5; DB 1; Length 2894;			
Best local similarity 19.0%; Pred. No. 0.0027;			
Matches 383; Conservative 250; Mismatches 719; Indels 659; Gaps 105;			
QY 65 LSYSPLNKTDIGFGIGFGLSVYDRKNSLSLSTGKNGYKVIETDKTVKLOQKLDNLP 124			



Db 409 ISYDPVNEKEI---IKLEGESEKIKLLRYEKKKYKI---SKIVKIYNSVIBIKL 460  
QY 125 EKD---LKE---NCYRIIHKSGDIEVLTFNNNA---FDLKV---KLL 162  
Db 461 KADREILLKYNLPENILNTTEIVKISNKIRVEVNNKVDGWRFSCKIPKGRVKEIV 520  
QY 163 NPAGHAI---YIDNPEATQPLNRIV---DLDGDHIDPLL--- 197  
Db 521 GDDGRVIRNNISINRLTEGIVCEVRWVIE---NTLYFYDDPIYGYDISLIPPAPNHS 575  
QY 198 ---NLEVOGLIKITLTFEQKEGYETELRFLN---RQLNSIHNFSLGNE 241  
Db 576 IAVELSYNGQYGGCGQISALVFPYKEDDETTVATYDHAGRTGDYVANNIDAYA-GSK 634  
QY 242 NPLTWSFG---YTPIGKNGILQW---ITSMTPAGGLKETVNS--- 279  
Db 635 IAIKYTSGALRQYGVLTAGSLGWWTYYLSEINRDTPLNTVPNGILESVIITDMYAP 694  
QY 280 ---NNQGHFFPOSANLPVLYTLTKQVPGAGQALQAEY---SYTSHNYVGG--- 327  
Db 695 WNNNELN---ITQKVIRGNKWFATYIYKNTPTKTYTLNLFQGGMDWN 741  
QY 328 ---GSGNIGWNNKLDNLYGLMTEYNYGSTESRYKDKHGDQIVRIERTYNNYHLLTS 381  
Db 742 FRGSWGGDAYNSIDDVWYGDENAPVGDIOYGGFK---SNIPSEHDVNLWYSTWS 796  
QY 382 ECKOON---GYIQTETA---YVAI---IGHNFSQPSQF---QL 414  
Db 797 DIRYDNLNDSYEGDAGTALAWTKSLKPGEIWVPIIWMGLGYNTDMNNEINMGLSQL 856  
QY 415 PKT---KTETWSADNSRSEI---TETTFDESIGN 443  
Db 857 YDTGVKSIDYFNNGSDSNPIIPIIYINSITIALYGLVDAYNLVNSINITQINGTYITNS 916  
QY 444 PLTKVIKDKTKQKIIISPESTHWEYYPAGEVDNCPBPYGFTRFKKIIQT---PYDSEFKD 501  
Db 917 TLINLSVPYEBEKLVS---FPV---NISNMEPYGAYNIT---IKTNLPNDQNTSN 961  
QY 502 DPERFIQY---RYSLIGSOSHVTLKIEBHYHSATOLLNSTLFOYNTDKSELGR-LKOTEC 558  
Db 962 DEKILIIYISFSVOPVQKTNVGEIP---YNTILYNFVG---GRDINITYL 1012  
QY 559 TKGENGKTY---VVKFTYTKQDDTLQOQSHSIITHDNFTHRSQVR 602  
Db 1013 TKGWTTKIYNSILIAEDANGDGDWIDVNP-NYDLNSNNLPDIIVPTGEINLTVSKTIPS 1071  
QY 603 SRYTGRLEFSDTDKDIVTQMSYDKLGRLLTRTNSGTPYANTLTIDYELN---NLQDDN 658  
Db 1072 TAPLGEI---DTTILKFVNNINPSIFGK---TTFQTSTPYPPSVQKTFYLGHTLRLTNTS 1126  
QY 659 RPPFVITTTDVGNGQL---RNEFDGAGRHSQCCLKSDGDGKFYTIHTQQYDEQ 709  
Db 1127 IPTTINNYTINSLSASWIOYPRFADNFTVVGK-IPILLYINDPNVIFGT--- 1176  
QY 710 GRHTSTVSYDLTNGROQ---TDPDKVHLSMS-KSYDNMGQIANTHWSYGVSEKITTVP 764  
Db 1177 -EMEKIVVSLMATNGIDSFTLGSDBEYLYLDDTIKSY---IFN---ITLDS 1220  
QY 765 I---TLTATKOLQSNVNOVQKEVTTVPSCQPIQITLPEAGHLOSCHT--L 813  
Db 1221 IITIPKNYLVLRVENQIISNSINIEY---STY-PSNIYNTITTYVNVNIFSDKNVYL 1276  
QY 814 TRDGWDRVRKETDAIGCTIYQYDYNVRVQITLPDGTI-VNR-----KVAP----- 859  
Db 1277 PNEVNTIFANITDPIG---SYDISGANITVYVYNGSVVINSMLLQEIIDKXNSPSLWL 1331  
QY 860 ---FSTD---TLTIDIRVNGI-----SLGQOFDGLSRLTQSDGG 894  
Db 1332 YNVSFSLPESKYLITITIGIESNGVSKKNYTYCGYIOGYKVEDFTLGR-EDSEDKG 1390  
QY 895 RVMAYTYS---AGNDQCP---STVI---TPDGOFIHYQ---YQPELDDAVL----- 933

Db 1391 -IYGVNVSLLENNNDGIPDIDGTIVNSTTTDIFGHYSFLVYNSKTYFVVVNSRTVGT 1449  
QY 934 ---OVASNEI---TQOFSYNPVTGALLKVAEQS-----LTPYIYPSGRL-- 973  
Db 1450 RGLNPOYSKNDIWAETIYQVTPINSS---QWIANGNASIFPDKLLITDDTGEAGSVWY 1507  
QY 974 -XWENINDMKMSYMLWTLRGLNGYDITLGTIQTISRD-----HGRVTQIKDSSI 1023  
Db 1508 YKPVNLSEDLVVEFYAYLGDNDPDGADITFTLQSLGTNELGGTGBLGGISPSVAVEV 1567  
QY 1024 KTTLN-YD-----DLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGKLCUSSGH 1072  
Db 1568 DTWLNDFDAPATDTHAIDVNGINHTYNSLT---YPTPNYDILGNVEDGRE-----H 1617  
QY 1073 TLDIOQSWLKTQOLANRIVKLVNGVLTQRTBOYSYDSNRNLNOYKCDGAECPDKYHSIYV 1132  
Db 1618 LIKI---VWNAI-----TKLQVYFDGNLSLTWNK-----DIT 1647  
QY 1133 QNFYDIYINITACHTTPADGTEDHATFKFANPT-----DPCQLTEVHT-----HPD 1180  
Db 1648 Q-----ILGN---SAYPGFTGGTGGAKNLOQYVPIVYKNGDGYIINPTYGVMFEGORDN 1700  
QY 1181 MPDNIR-LKYDKAGRVINITDNHGN---TENFTYDILGRQN-GQGS-----VYGY 1226  
Db 1701 EEDNWEDGKYEHY-CLINLNSYSYGNITFGPSFDVITNTKSTGQSFQFIKNANAIYK 1759  
QY 1227 DPLNRLVSOKTDLCELYRETMLVNEVRNGEMIRLLRTGETII-----AQRASK 1278  
Db 1760 D-----ESYFRIPNI---DARKNGHY-IVTSGNKILDLNLTIVNGSTQINGT 1801  
QY 1279 VLLTG-----TDSQOSVILTSDKQNLQSOEAYSAYGKHKSTANDASI-----LG 1321  
Db 1802 IILSGLOWTANGVAYNNNSNLTILTPDDYN---QKGSVWYKYPVNLSEDLVVEFYAYLG 1859  
QY 1322 YNGERADPV---GVTHLGN---GYRSYDPT-----LMRHTP---DLSLSPF 1359  
Db 1860 DNPDCADGITFTLQSLGTNELGGTGLGYGSPSVAVEVDTLNDFDSPATTDHAIAD 1919  
QY 1360 GAGGIN-----PYSYCLGDPINRSDPSGHLISWQAWTGTGIGWGIAGLLLTATGGWA 1409  
Db 1920 VQGNLNTYNSITYPNPYDLG---NVEDGREHLIKIVWN----- 1957  
QY 1410 IAAAGGIAAAIATSTTALAFG---ALSVTSDITSIVSGALEDASPKASSILGWSMGMG 1466  
Db 1958 ---ATTKTLOVYFDGNLALTNKDIQIGN-----STYFGTGGTGG 1997  
QY 1467 AAGL---ABSAIKGKTGLATHLGAFAEDGENALLKSTSESEIKWGVTRSLDREIVRNEE 1523  
Db 1998 AKNLQVVKPYVYKNGDNLN---LEEISPNPIIDNVGADTYIGNIFFENVSUGILGNET 2053  
QY 1524 G-----QVIKDHSG---YTDNFM---GKGEQAILVHGDKGFL 1556  
Db 2054 GLNNLTLSKSGIYKILNAGVKLVYDWSLQNYPIYIDLNTINASGGYIGISMLNKIWML 2113  
QY 1557 YHTEGNKNGKOPY-----TRHTPEQLVDYLKDNVIVDLTQGGDKPVHL 1600  
Db 2114 YNSQISLKNGVGIYWANWAGFGNITTYNITSSNOGLVLYKDGNGIKLINSQIKNSVY 2173  
QY 1601 LSCYKSGSAAADKMAKYNRPVIAYSNKPTI 1631  
Db 2174 EGYVSKNSTLEILNSSIINNSIGIYANISSI 2204

RESULT 8  
P2P LACLC  
ID P2P LACLC STANDARD; PRT; 1902 AA.  
AC P15293;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
DE associated serine proteinase) (LP151).  
GN PRT.



Db 1304 NSVTFDQGVTFGANEFNATSAKFDYDPKGTIATITGKVKHPTTLTQVQDKQIPI-----KDD 1359  
QY 1131 VTQNTFYD-----LYGNITACHT-----TF-----ADGTEHATKRFAN 1164  
Db 1360 LTFSTLTLGLGQKPGVGVGDTQNTKTFQEAFTILDVAFTLSLDSSTDPVIT-NW 1418  
QY 1165 P-----TDCQLTEVH-----HTHDPMDPNIRLYDKAGRV 1195  
Db 1419 FNQITGTATDQAQYLSLSINGSSVASQYVDININSKPGHMAIDQP--VKLLEGKNVLT 1476  
QY 1196 INITDNEGN--TENFTYDTLGLQNGQGVYGYDPLNRLVYSQKTDTLDCELYRETMVYN 1253  
Db 1477 VAVTDSNNTTKMIT-----VY-IEPKTILAPVTPSTTTPAKTVITLTA 1522  
QY 1254 EVRNGEMIR-----LLRGTETIIA-----QQRASKVLLTGTDSOQ 1288  
Db 1523 SAATGETVQVSADGGKTYQDVPAAGVTVTANGTFKPKSTDLYGNEPVDVYVVTNIKADD 1582  
QY 1289 SVILTSKONL-----SQEAYSACKH-----KSTAND-----ASILYNGER 1326  
Db 1583 PAQLQRAKQELNLIASAKTSLASGKYDDATTALAAATOKAQOTALDQTNASVDSUTGAN 1642  
QY 1327 AD-----PVSQVTHLNGYRSYDPTLMRFHTPDSLPFGAGGINPYSCIGLDP 1374  
Db 1643 RDLQTAINQLAALKLPADKKTSLNLQLOSVKAALETDLGNQTDSTG-----KTFEALDDL 1698  
QY 1375 INRSDPSGHLISQWAWTGIGMGIAGLLTIATG-----GMAIAAGGIA--AIAAST 1423  
Db 1699 VAQAQAGTQDDQQLQATLAKVLDLAVLAKLAEGIKAAATPAEVGNAKDAATKTYADIAT 1758  
QY 1424 STTALA-----FGALSVTSDDITSIVSGALEDASPKASSILGWSMGMAAGLAESAIAK 1476  
Db 1759 LTSQASADASDKLAHLQALQSLKTKYVAAVEAAKT-----VGKGDSTTTSKGGG 1810  
QY 1477 GGTKLATHGAFBDGNALLKSTSESRIKGWVTRSLREIVRNBEGQVHKDHSRGYTD 1536  
Db 1811 QGTPAPTPGIGDKGDEGQSPSGGNIPTNPAITTSSTDDTDRNGQLTS-----1862  
QY 1537 NFMKGK 1542  
Db 1863 ---GKG 1865

## RESULT 9

P3P\_LACLC STANDARD; PRT; 1902 AA.  
AC P15292; 1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE P111-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
associated serine proteinase).  
GN PRTP.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RS SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.  
RC STRAIN=SK11.  
RX MEDLINE=89340435; PubMed=2760036;  
RA Vos P., Simons G., Slezien R.J., de Vos W.M.;  
RT "Primary structure and organization of the gene for a procaryotic,  
cell envelope-located serine proteinase.";  
RL J. Biol. Chem. 264:13579-13585(1989).  
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
GROWTH OF THE BACTERIA ON MILK.  
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
specificity, although some substrate preference have been noted,  
e.g. large hydrophobic residues in the P1 and P4 positions, and  
Pro in the P2 position. Best known for its action on caseins,  
although it has been shown to hydrolyze hemoglobin and oxidized

CC insulin B-chain.  
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -1- SIMILARITY: Belongs to peptidase family S8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; J04962; AAA03533.1; ALT\_SEQ.  
DR HSP; P00782; 2SBT.  
DR MEROPS; S08.019; --  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;  
KW Signal; Plasmid.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1870  
FT PROPEP 1871 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT SITE 1867 1871  
FT MOD\_RES 1870 1870  
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;  
FT  
FT P111-TYPE PROTEINASE.  
FT REMOVED BY SORTASE (POTENTIAL).  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT LPXTG SORTING SIGNAL (POTENTIAL).  
FT AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ  
Query Match 2.3%; Score 205.5; DB 1; Length 1902;  
Best Local Similarity 18.7%; Pred. No. 0.0055;  
Matches 316; Conservative 195; Mismatches 589; Indels 586; Gaps 83;

QY 239 GNEPLT---WSFGYTPIGKNGILQ---WITSATPG-GLKETV-----NYSNNQ- 284  
Db 521 GTATPMTSIALTTPTFTGLSVTGQKLVDWVTAHPDSDSLGKVTLAMLPNQKYTEDKMS 580  
QY 285 H--HFPQSANLPVLPVTLMKQVPGAGQPAQAEVSYTSHNYVG-----GSGNGIMNKK 336  
Db 581 DFTSYGVPVSNLSFKPDIT---APGNTWSTQNNNGYTNMGSTMASPFIAGSQALLKQA 636  
QY 337 LDNLXGLMTEYNGSTESRRYKDEG---HQIVRIER-----TNNYHLLTSECK 384  
Db 637 LNNKNPPFYA-----YQKLGATLDTFLKVTENMTAQPINNINN--VIVSPRR 685  
QY 385 QQNGYIQ-----TTETAYVAILIIGHNFDSPQSFQLPKTKTETWRSADNSY 429  
Db 686 QGAGLVYDKAAIDALEKNPSTVVAENGYPVELKDFSTDKTKLTFNTRTHELTYQMD 745  
QY 430 RSEITETTFDESNGNPLTKVTKDKTKIIPSTHWEYYPAGEVNDCEPE-----PYG 482  
Db 746 SNTDTNAVYTSATDPNSGLYDKK-----IDGAAKAGSNITVPAG 786  
QY 483 FTRFVKKLIQPYDSEFKDDSEKFIQYRYSILGQSHVTLXIEERHYSATOLLNSTLPQY 542  
Db 787 KTAQTEFTLSLPKSPDQOQFVEGLNF-----813  
QY 543 NTDKSELGRLLKQTECKGKNGKTYVSVHKFTYTKQDDTLQOQSHITTHDNFTIHRQVR 602  
Db 814 -----KSGDG-----SRLN 822

QY 603 SRYTRFLPSDDTDKDIYVQM---SYDKLGR-----LLTRTLNSGTYPANTLTIDYELNLL 654  
Db 823 LPYMG-FFGWNDGKIVDSLNGITYSPAGNFGFVPLLNKNTGTQYGGMVTADGNKT 881  
QY 655 QDDNRPPFVITTDVNG-NOLRNEF-----DGAGRHVSQCLKSDSDGDKF 698  
Db 882 VDDQAIAF---SSDKNALYNDISMKYVLLBNISNVQVDILDGQGNKVTLLSSSTNEKKY 938  
QY 699 YTIHQ-----YDEQGRHHTSYSD---YLNG-----RQOTD----- 729  
Db 939 YNAHSQQVYYNAPADGTYDQDGNIKTADDSYTYRISGVPEGGDKRQVDFVPKLD 998  
QY 730 ---PDKHLMSKSYDNWQLANTHWSYGVSEKITVDPIITLTATKOLQSNVNVQTKEV 786  
Db 999 SKAPTVRHVALSAKTEN-GKT-----QYLTAEAKDLSGLDATKSVKTEINEV-TNLDA 1051  
QY 787 TTYTPSQOPIQITLDEAGHLQSCHTLTRDGDWRV-----KETDAIG-----QCTIYQYD 837  
Db 1052 T-----FTDAG-----TTADGYTKIETPLSDEQAQALNGDNSAELYLTD 1091  
QY 838 NYNRVIQITLPDGTIVNRKYPAPSTDITLITRVNGISLGCQTFDGLSRLT-----QSOD 892  
Db 1092 NASNA---TDQASV---QKPGSTs-----FDLIVNGGIP-----DKISSTTGYEANTQG 1137  
QY 893 GGRVMAVYSAGNDQCPSTVITPDGQFIHYQYQPELDQVAVASNEITQOFSYNPVTA 952  
Db 1138 GG---TYTFSG---TYPAV---DGTYDAQGGKXHLNLTYYDAATNSFTASV---PVTNA 1185  
QY 953 LKVAEGQSUTPIYPSGRUKMENINDMKOMSVLWTLRGLNGYDTLTGTQIKSRDTH 1012  
Db 1186 DYAAQVD-----LYADKAHTQLLKHFDTK-----VRLMAPTFDL-----KFNGNS- 1226  
QY 1013 GRVTOIKDSSIKT---TLNYDNLNHHIGSQVTDLATGMLTTFVDFGLNREIRKGLCDSS 1070  
Db 1227 ---DOTSEATIKVGTVSADTKTVNGHTVAALDAQHHFSVDPVNVYGDNTIKVATDKD 1283  
QY 1071 GHTLDIQO---SWLKTQOLANRIVKLVGLVORTQYS-----YDSNRRL----- 1111  
Db 1284 GNTTEQKTISSYDPPDKKSVTFDQGVKFTKFNATSAKFPDKTGTATITGKVKHP 1343  
QY 1112 -NQVKDCACBPTKYGHSIVTQFTYDI-----YGNi-----TACHTTADGTE----- 1155  
Db 1344 TTTLOVDGKQPI-----KDDLTFTFTLDLGLGKPGVGVVGDITQNTQEQALSFTLDA 1399  
QY 1156 -----DHATFKFANPTDP-----CQTEVH-----H 1176  
Db 1400 VAPTLSDSSDAPVYNDPNFQITGATDNAQVLSLSINGSSVASQVEDININSGKPGH 1459  
QY 1177 THPDVFNIRLKYKAGRVINITDNGN---TENETDYLGRLOQGSVYGYDPLNRLVS 1234  
Db 1460 MAIDQP---VKLEGKQVLTAVTQSDENNTTKNIT-----VY-YEPKKTAA 1503  
QY 1235 QKTDLDCELYRETMVNEVNGEMIR-----LLRTGETIIA----- 1272  
Db 1504 PTVPSTTEPAQTVTLTANAATGETVQYSADGKTYQDVPAAGVITANGTFKFXSTDL 1563  
QY 1273 ---QORASKVLLTGTDSQSVILSDKNL-----SQEAYSAYGH----- 1310  
Db 1564 YGNESPAVDYVYVNTIKADPPAQQAQKQELNLIASAKTUSASQKYYDDATTALAAATQK 1623  
QY 1311 KSTAND---ASILGVNGERAD-----PVSGVTHLNGYRSYDPTFLMRFHTPDS 1355  
Db 1624 ACTALDQTNASVDSLTGANRDLQTAINQAAKLPAKDKTSLNLQOS-----VKDA 1674  
QY 1356 LSPFGAGGINPYSYCLGDPINRSDPSGHLSSQAWTGICMGAGIAGLLLIATGGMIAAAGG 1415  
Db 1675 LGT-----DLGNQTDPS-----TGKTFTAALDQLVAQQAQ 1705  
QY 1416 -----IAAAIATSTTALAFGLSVTSDITSIVSGALEDASPKA---SSILGWVSMGMA 1467  
Db 1706 TOTDQLOQATLAKILDEVLAKLAEGIKAAATPAEVGNAKDAATGKTWADIADTLTSGAS 1765  
QY 1468 AGLAESAIKGGTKLATHI-GAFAEDGENALLKSTSESSRIKGVTRSLDRIVREBEGQVI 1527

Db 1766 ADASD-----KLA-HLQAL-----QSLKTKVAAAEAAKTVGKG----- 1798  
QY 1528 KDHSGYTDNFMKGEGEQAILVHGD---KDGFLYHTEGNKHNGK-----PYTRHTPEQLVD 1580  
Db 1799 -DGTITGTDKGGGGQTPA-PAPGDTGKD---KDEGSGQPSGGNIPTKPAF---TTSPTTD 1851  
QY 1581 YLKNNNIVDLTQGGDKPVHLLSCYKSGSAAADKMAKVINRVIAYSNKPTIS---OCLA 1636  
Db 1852 DTTDRN-QOLTSQ-----XGALPKTGETTERPAFGFLGVIVSLMGLGLK 1896  
QY 1637 RIERKD 1642  
Db 1897 RKOREE 1902  
RESULT 10  
PIP LACLC  
ID PIP LACLC STANDARD; PRT; 1902 AA.  
AC P16271;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE P1-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine  
proteinase).  
GN PRTP.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid pW05.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-WG2;  
RC MEDLINE=88149035; PubMed=3278687;  
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledeboer A.M., Venema G.;  
RT "Nucleotide sequence of the cell wall proteinase gene of  
Streptococcus cremoris Wg2";  
RL Appl. Environ. Microbiol. 54:231-238 (1988).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
GROWTH OF THE BACTERIA ON MILK.  
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
specificity, although some substrate preference have been noted,  
e.g. large hydrophobic residues in the P1 and P4 positions, and  
Pro in the P2 position. Best known for its action on caseins,  
although it has been shown to hydrolyze hemoglobin and oxidized  
insulin B-chain.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- SIMILARITY: Belongs to peptidase family S8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M24767; AAA17677.1; --  
DR HSSP; P00782; L501.  
DR MEROPS; S08.019; --  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00882; Peptidase\_S8; 1.  
DR PRINTS; PRO0723; SUBTILASIN.  
DR TIGRFAMs; TIGR01167; LPX1G\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.

[illegible]

RC STRAIN=NCDO 151;  
RX MEDLINE=92381481; PubMed=1512565;  
RA Hoick A., Naes H.;  
RT "Cloning, sequencing and expression of the gene encoding the cell-  
RT envelope-associated proteinase from *Lactobacillus paracasei* subsp.  
RT paracasei NCDO 151";  
RL J. Gen. Microbiol. 138:1353-1364 (1992).  
RN [2]  
RP SEQUENCE OF 189-196.  
RX MEDLINE=92226694; PubMed=1564442;  
RA Naes H., Nissen-Meyer J.;  
RT "Purification and N-terminal amino acid sequence determination of the  
RT cell-wall-bound proteinase from *Lactobacillus paracasei* subsp.  
RT paracasei";  
RL J. Gen. Microbiol. 138:313-318 (1992).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
CC specificity, although some substrate preference have been noted,  
CC e.g. large hydrophobic residues in the P1 and P4 positions, and  
CC Pro in the P2 position. Best known for its action on caseins.  
CC although it has been shown to hydrolyze hemoglobin and oxidized  
CC insulin B-chain.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
CC an amide bond (Potential).  
CC -!- SIMILARITY: Belongs to peptidase family S8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M83946; AAA25248.1; -;  
DR PIR; B44858; B44858.  
DR HSP; P00782; I801.  
DR MEROPS; S08.019; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR001337; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPTXG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;  
KW Signal.  
FT SIGNAL. 1 33 POTENTIAL.  
FT PROPEP 34 187 POTENTIAL.  
FT CHAIN 188 1870 PII-TYPE PROTEINASE.  
FT PROPEP 1871 1902 REMOVED BY SORTASE (POTENTIAL).  
FT ACT\_SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT MOD\_RES 1867 1871 LPTXG\_SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 1902 AA; 200253 MW; D8C9F38CE5DA582 CRC64;  
  
Query Match 2.3%; Score 200.5; DB 1; Length 1902;  
Best Local Similarity 19.2%; Pred. No. 0.01; Mismatches 659; Indels 645; Gaps 94;  
Matches 361; Conservative 215;  
  
Y 12 ITMSDNEFFQANNFTSAVSGVDPRPTGLYNIQITLGHIVGNGLGPTLPLTSLSPN 71  
D 419 VASAENTVISQAVTIID-----GKDLQIGPETIQLSSNDFTG----- 457  
Y 72 KTDIGFIGNFGLSVDRKN-SLLSLSTGENYKVIETDKTVLQXK-----KLDNLRPEK 126

Db 458 -----FDQKFYVVKDASGDLKGAADYTADAKGKIAIVKRGELAFAD 501  
Q 127 DLKENCYR-----IIHSG-----DIEVTGFNNNAFDLKVPKKLNPAAGHAIYI 171  
D 502 KQK-----YAAAGAAAGLIIVNNDGTATPLTSLRTTPTFGLSSKTQGL-----V 549  
Q 172 DWNFEATOPRLNRIYDDLGDHDIPLNLEYOGLKLTILPFGQKEGVRTELRLNQLN 231  
D 550 DW--VTAHP-----DSSLG-----VKIALTLLPNQY----- 574  
Q 232 SIHNFSLGNENPLTWSFGYTPIGKNGILGOWITSMATPGGLKETVYNNQGHFFQSA 291  
D 575 -----TEDKMSDFTSYGPVSNL-----SPKPDITAPGG-----NIWSTQNNNG-YTNMSG 618  
Q 292 NLPVLPYVTLMKQVPGAGQPAIQAEYSYTSYHNYGGSGNGIWNKLDNLXGLMTEYNYGS 351  
D 619 TSMASFFI-----AGSQALLKQ-----ALNNKNPPFYA----- 646  
Q 352 TESRRYKDKEG---HDQIVRIER-----TYNNYHLTSECKQNGYIQ----- 391  
D 647 ---DYKQLKGALTDLFKTVMNTAQPIINDINYN--VIVSPRQAGLVDVKAADAL 700  
Q 392 -----TTETAYVAIIGHNFDSPSQPOLPKTKETWRSADNSYRSEIETTFDESGNP 444  
D 701 EKNPSTVVAENGYPFAVELKDFSTDKTKFLTTRTHELTYQMSDNTDNAVVTSATDP 760  
Q 445 LTKVIKDKKTO---KIISPTHWYYPAGEVNDPCPEYGFTRFVKKIIQTDPYSEFK 500  
D 761 NSGVLYDKKIDGAAIKAGSDIT-----VPAGKTAQIEFTLSLPKSDQ 804  
Q 501 DDEPKFIQYRSLGQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLKQTECTK 560  
D 805 QFVEGFLNFKGS-DGSR-----LNLPMGFFGDMND-GKIVDSL----- 841  
Q 561 GENGKTYSVVHKFTYTKQDDTLQOQSHSITTHDNFTIHRQSVRSYTGRLFSDDTKDIVT 620  
D 842 ---NGITYSPAGNGYGV-----PLLTWN-TGH-----QYGGWTDADGKTVD 883  
Q 621 Q-----MSYDKLGRLLTRTLNSGTPYANLTVD-----YELNNQDNRPPVIITDUNG 671  
D 884 DQAIASFSDK-----NALYNDISQWYLLNLSN-----VQVDI-- 917  
Q 672 NQLRNEFDGAGRHYSOCLKSDGDKFVTHIQ-----YDEQGRHSTYSVD 719  
D 918 -----LDGQGNKVTTLSSTNTQTYDAHSQKIYINAPAWDGTYYDQDGNIKTADD 971  
Q 720 ---YLTNG-----ROOTD-----PDKVHLSMSKSYDNWGOIANTHWSYGVSEK 759  
D 972 GSYTVRIISGVEGGDKRQVDFVPFKLDSKAPTVAHVVALSAKTEN-GKT-----QYLTAE 1025  
Q 760 ITVDPIITATKQIQSNNSNNVTGKENTYTFPSQOPIQITLFDAGHLQSCHTITRQWD 819  
D 1026 AKODLSGLDATSKVKTAINEV-TNLDAT-----FTDAG-----TTADGYT 1064  
Q 820 RVR-----KETDAIG-----OCTIYQYDNNYRVIQITLPDGTIVNRKYAPSTDTLITDIR 870  
D 1065 KIETPLSDEQAQALGNGNSAELVITDNASNA--TNQDASV-----QKPGSTS--FDLI 1114  
Q 871 VNGISLGQOQTFDGLSRLT-----SQDQGRVWATYSAGNQDQCSVTITPDGQIHVYQY 925  
D 1115 VNGGIP-----DKISSITTYEANTQGGG---TYTFSG---TYFAAV---DGTITDAGK 1161  
Q 926 PELDDAVLQVASNEITQQFSYNPVTGALLKAVABGQSITPIYPSGRKXENINDMKMS 985  
D 1162 KHDLTNTTYDAATNSFTASWA---VTNADYAAQVD-----LYADKAHTQLLKHFDTK--- 1209  
Q 986 YLWTLRGLNENGYDITGTIQIKISRDTHGRVTVQIKDSSIKT---TLNYDDLNRHIGSOYD 1043  
D 1210 ---VRLTAPTFTDL-----KFNNGS-----DOTSEATIKVTGTVSSDPTKTVNGDVAAL 1256  
Q 1044 ATGHMLTITVTFDGLNRIEGRKLCDSGHTLIDQSSMLKTCQLANRIVKLNGVLQRTQEQ 1103

Db 1257 DAGHFDVDFVNVGDNITKVTATDEDNNTTBEQKITSSYD-----PDVLKNVTF 1308  
QY 1104 SYDSRNELNOYKCDGACPTDKYGHSTVTONFTYDIYGNITACHTTF-ADGTE----DHA 1158  
Db 1309 DQGVKFCANFENATSAKFPDKGIATIT-----GKVKHPTTTLQVDGKQISIKNDL 1360  
QY 1159 TFKFANPTDCQLTEVHHPTDMPDNLRLKYDKAGRVINITDNHGNENTFYD-----1211  
Db 1361 TFSPT-----LDLGLTGKQKFPVVVGDT-----TONKTFQALTFIL 1397  
QY 1212 -----TLGLRQNGQGVYGDPLNRLVSQKTDITLDCELYRRETMVNEVNGEMIRLLRT 1266  
Db 1398 DAVAPTLSSSTDPAPVYNDPFOITGTAID-----NAQVLSLAIN 1439  
QY 1267 GETIAQCRASKV-----LITGDSQOSVILTSDKQNLSEAYSGKH 1311  
Db 1440 GSHVASQYADININSGKPGHMAIDQPVKLEGG-KNVLTVAVTSDSENNTTKITVYEPK 1498  
QY 1312 S-----TANDASILGNCE-----RAD-----PVSGVTHLNG 1339  
Db 1499 KTLAAPTFTSTTEPAKTVTLTANAAA-----TGTEVOYSADGGKTQVDVPAAGVTVTANG 1554  
QY 1340 YRSYDPTLMPFHTPDSLPFGAGINPYSYCLGDPINRSPSG-HLSWQAWTIGMGIAG 1398  
Db 1555 -----TFKXSTDLV-----GNESPAVDVVTNIKADDPQAOLQAKQALTNL-----IAS 1599  
QY 1399 LLLTIATGMAIAAGGIAAIASTSTTALAFGALSVTSDITSIVSGALED-----1449  
Db 1600 AKTLSAGSKYDDATTTALAAATQKAQT-----ALDQTDASVDSLGTANEDLOTAINQLA 1653  
QY 1450 ---ASPKASSITLGVSMGMGAAG-----LAESAIGKGTKLATHLGAFAEDGENALLKST 1500  
Db 1654 AKLPADKKTSLNLOLQSVKALGTDLGNQDTPSTGKFTTAAALDVLVAQAAG-----T 1706  
QY 1501 SESSRIKGVTRSJDRBIVRNEEG-----QV---IKHNSRG---YTD---NFKMGKEQAI 1546  
Db 1707 QTADQLQASLAKVLDAVLAKLAEGIKAAKTAPEVGNNAKDAATGKTWYADIADTLTSGQAS- 1765  
QY 1547 LVHGDGKGFYLVH-----TEGKNHKGKGFYTRHTPEO-----LVDYL 1582  
Db 1766 ---ADASDKLAHLQALQSLKTKVAVAEAAKTAGKDDTTGTSKGGGGQTPAPAGDTG 1822  
QY 1583 KD---NNIVDLTQGGD---KPVHLLSC-----YKSSGAADKQAKYINRPVIAYS 1626  
Db 1823 KDKGDEGQSSGGNIFPKPATITSTSTDDTTDANGQHTSGKGALEKPTAETTERPAPGFL 1882  
QY 1627 NKPTIS-----QGLARIERKD 1642  
Db 1883 GVIVYSLMGLGLKQKQREE 1902

## RESULT 12

ID YEEJ\_ECO57 STANDARD; PRT: 2660 AA.  
AC Q8X8V7; Q8X2B9; Q8X2C0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein yeeJ.  
GN 23135 OR\_ECS2775/ECS2776.  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11208551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
RN Nature 409:529-533 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohsuho E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:11-22 (2001).  
CC -!- SIMILARITY: Contains 16 Big-1 domains.  
CC -!- SIMILARITY: Belongs to the intimin/invasin family.  
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a  
CC frameshift in position 1315.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
CC EMBL; AF005423; AAC57041.1; -----  
DR EMBL; AF002559; BAB36198.1; ALT\_FRAME.  
DR EMBL; AF002559; BAB36199.1; ALT\_FRAME.  
DR InterPro; IPR003344; Big 1.  
DR InterPro; IPR003535; Intimin.  
DR InterPro; IPR008964; Invasin\_intimin.  
DR InterPro; IPR000601; PKD.  
DR Pfam; PF02369; Big\_1; 16.  
DR PRINTS; PR01369; INTIMIN.  
DR SMART; SM00634; BID\_1; 16.  
DR SMART; SM00089; PKD; 8.  
DR KW Hypothetical protein; Repeat; Complete proteome.  
FT DOMAIN 738 834  
FT BIG-1 1.  
FT DOMAIN 840 929  
FT BIG-1 2.  
FT DOMAIN 931 1033  
FT BIG-1 3.  
FT DOMAIN 1042 1132  
FT BIG-1 4.  
FT DOMAIN 1134 1236  
FT BIG-1 5.  
FT DOMAIN 1245 1335  
FT BIG-1 6.  
FT DOMAIN 1337 1439  
FT BIG-1 7.  
FT DOMAIN 1448 1539  
FT BIG-1 8.  
FT DOMAIN 1548 1652  
FT BIG-1 9.  
FT DOMAIN 1653 1750  
FT BIG-1 10.  
FT DOMAIN 1751 1855  
FT BIG-1 11.  
FT DOMAIN 1856 1957  
FT BIG-1 12.  
FT DOMAIN 1963 2056  
FT BIG-1 13.  
FT DOMAIN 2065 2156  
FT BIG-1 14.  
FT DOMAIN 2157 2252  
FT BIG-1 15.  
FT DOMAIN 2254 2355  
FT BIG-1 16.  
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;  
Query Match 2.2%; Score 198; DB 1; Length 2660;  
Best Local Similarity 18.7%; Pred. No. 0.024;  
Matches 288; Conservative 187; Mismatches 580; Indels 486; Gaps 67;  
QY 228 RQINSHNPSLGNPLTWSFGYTPGK---NGILGQWITSMTPAGLKETVYNSNNG 284  
Db 671 QQLNTA--VSDNVKRP-----GVTTDWKETADGVYKATYATYKSGGLTAKLMMWNED 723  
QY 285 HHF-----POSANLPVLPVYVTLMKQVPGAPQPAIOAEVSYTSHNYVGGSGNGIWNK 336  
Db 724 LHTAGFIDANPQSAKIATL-----SASNNGLANANAATVSVNVADEGSFNINDHT 776  
QY 337 LD--NLYGLMTEYNGSTESRRYKDEGHDQIVRIERTYNNVHLLTSECKQONGYIOTTE 394  
Db 777 VTFPAVLGSGATSFNQNTAKTDVNGLATFD--LKSSKQEDN----TVEVTLNENGVKQT-- 828

395 TAYYAIIGHNEDSOPQOLPKTKETWRSADNSYRSEITETTTDESNGPLTKVICKDKT 454  
829 ---LIVSFVGSSTAQVDLQSKNEV---VADGNSATMTATVRDAKGNLLNDV---KVT 879  
455 QKIISPSHWEYYPAGEVNDPCPEPYGFTFRVKKIIQTTPYDSEFFKDDPEKFIQYRSLI 514  
880 FNVNSAAA-----KLSQTEVNSH-----D 898  
515 GSQSHVTLKIBERHYSATOLLNSTLFQYNTDKSELGRLLKQTECTKENGK---TYSVVH 571  
899 GIATATLTSKNGDYTVTASVSS-----GSQANQOQVIFIGDQSTAALTLSV-- 944  
572 KFTYTKQDDTLQOQSHSIITHNFTHRSQVRSRYTGRFSDTD-----TKDITVQMSYDK 626  
945 -----PSGDIIVTWTAPLHMTATLQDKNGNPLKDKKEITFSPVPNDVASRFSISN 992  
627 LGRLLT-----RTNSGTPYANTLTYDYELNLIQDNRPPFVI 664  
993 SGKGMDSNGTAIASLTGTLAGTHMITARLANSNVSDTQPMTF-----VADKRAVVVL 1046  
665 TTT--DVNGN-----QLRNEFDGAGRHVSQCLKSDGDKGFYTIHTQQYDEQGRHH 713  
1047 QTSKAEIIGVDETTLTATVKDFDNNKLSVVFRTPSPADTQL--SLNARNTENGIAE 1105  
714 TS-----TYSDYLTNGRQOTD-----PDKVHLSMSKSYDNWGOQIANHWSYGVSEK 759  
1106 VTLKGTVLGVHTAEAILLNGNRDKIVNIAPDASNAQVTLNIPAQQVVTNNS----- 1157  
760 ITVDEILTATKQLOQSNVNVOTKCVYTYTPSQOPIQITLFDFA----- 804  
1158 ---DSVQLTAT--VKDPSNHPVAGITVNTFMPQDVAANFTLNNNGIAITQANGEAHVTLK 1212  
805 CHLOSCHTLT-----RDGMDRVKRETKDA-----IQOCTIY----- 834  
1213 GKAGHTVATLGNNSADAQPTFVADKDSAVVVLQTSKAEIIGNVDETTLTATVKD 1272  
835 QYDYNVRVIQTL---PDGIVNRKYAPFSTDILTIDIRVNGISLGOOTFDGSLRITQSQ 891  
1273 PFDNAVKDQVTFSTFNADTQLOSQSN--TNDSGVAEVTFGTGLVGVHTAEA----- 1323  
892 DGRVWYATYSAGNDQCPSTVITPDQGFIIHYQVPELDDAVLQASNE-----ITQ 943  
1324 -----TLPNGNNDKIVNIADAS---NAQVTLNIPAQQVVTNNSVSLTATVKD 1371  
944 PSYNPVTGALKKVAEQSLTPIYPSPGRUKMEN-----INDMKMSVLTILRGLNGYTD 999  
1372 PSNHPVAGITV-----NFTMPQDVAANFTLNNNGIAITQANGEAHV--TLKGGKAGTHT 1423  
1000 LTGTIQRISDRTHGRVTOIKDSSIKTTLNVDLNEHIGSOVTDLATGMLTITVTFPDGLN 1059  
1424 VTATLSNNNSDSQPTFVADKTSALVVLQISKEITGNGV--DSAT---LTATVK--DQFD 1478  
1060 REIGRK-----LCDSSGHTLDIQOSWLKTOQLANRIVKLVNGVORTQOYSYDSNRNLNQYK 1115  
1479 NEVNNLPVFTSTASSGLTLPGES--NTNESGIAQAATLQV----- 1517  
1116 CDGAECPDKYGHISVTONFYDNYGNIATCHTTFADGTEDHATFKFANPTDPCQTEVH 1175  
1518 -----APGETVITASIA-----NNGASDNKTVHIFIGDTAAKIIEL-- 1553  
1176 HTHPMDPNI---RLUKYDKAGRVINIT--DNH-----GNTENFTYDTLGRLONGQSGVYG 1226  
1554 ---TPVPDSIIAGTPQSSGSVITATVVDNNGFPVKGTVNTFNTSNAATAEMTNGQAVTN 1610  
1227 DPLNRLYSQKTDLDLCELYREIMLVNEVNGEMIRLLRGTETIIAQORASKVLLTGTD 1286  
1611 EQGKATVYNNTRSSIEGARPDTVEASLENGS--STLSTSINVNAD--ASTAHLTLLOA 1666  
1287 QOSVILTSKQNLQSOEAYSAYGKHKSSTANDASILNGVGERADPVSQVTHLNGYSYDPT 1346  
1667 LFDVTSAGDTNLLYIEVKDNY-----GNGVPPQOEVT 1697  
1347 LMRHTPDSLSFGAGAGGINPYSYCLGDPINRSDPSGHLSHQANTGIGMGIAGLLLIATG 1406

Db 1698 L-----SVSP--SEGTPSN-----NAYITNHGPN-YASFTATKAGVVQVATLENG 1743  
QY 1407 GMAIAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPASKASSILGWVSMGMG 1466  
Db 1744 D-----SMQQTIVVVPVNAEISLAASK----- 1767  
QY 1467 AAGLAESAALKGTGLATHLGAEDGENALLKSTSESSRIKMGVTRSLDRIVRN-----E 1522  
Db 1768 -----DPVIANNNDUTLTATVADTEGNAIANSE-----VTFILPDEVDRANFTLGD 1813  
QY 1523 EQQVIXDHSRGYTDNFMKGEGQAILVHGDQDGLYHT--EGNKHGKGPYTRHTPEQL--VD 1580  
Db 1814 GKGVTVD-----TEGRAKVTLKGTAG--AHTVTASMGKGS-----EQLVNVN 1854  
QY 1581 YLKD-----NNI-----VDLTQGGDKPV 1598  
Db 1855 FIADTLTAQVNLNVTFEDNFIANNVGMVTELOATVTDGNGNPL 1895  
RESULT 13  
RBP2 PLAVB  
ID RBP2 PLAVB STANDARD; PRT; 2867 AA.  
AC Q00799; Q9N2M3;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Reticulocyte binding protein 2 precursor (PVRBP-2).  
GN RBP-2 OR RBP2  
OS Plasmodium vivax (strain Belem).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=31273;  
RN [1]  
RP SEQUENCE FROM N.A., AND REVISIONS TO 2438-2439.  
RX MEDLINE=20299192; PubMed=10838229;  
RA Gallinski M.R., Xu M., Barnwell J.W.;  
RT "Plasmodium vivax reticulocyte binding protein-2 (PVRBP-2) shares structural features with PVRBP-1 and the Plasmodium yoelii 235 kDa rhoptry protein family";  
RT Mcl. Biochem. Parasitol. 108:257-262 (2000).  
RL [2]  
RN SEQUENCE OF 1189-2439 FROM N.A.  
RP MEDLINE=92315338; PubMed=1617731;  
RA Gallinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax merozoites";  
RL Cell 69:1213-1226 (1992).  
CC -!- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to human reticulocyte cells.  
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC OR send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC EMBL; AF184623; AAF76525.1; -.  
DR HSP; P03069; IGCN.  
KW Malaria; Receptor; Signal; Transmembrane; Repeat.  
FT SIGNAL 1 21  
FT CHAIN 22 2867 RETICULOCYTE BINDING PROTEIN 2.  
FT DOMAIN 22 2805 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 2806 2826 POTENTIAL.  
FT DOMAIN 2827 2867 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 44 133 ASN-RICH.  
FT DOMAIN 560 758 LYS-RICH.  
FT DOMAIN 1112 1285 7 X 4 AA TANDEM REPEATS OF H-D-D-T.  
FT REPEAT 2758 2761 1.  
FT REPEAT 2762 2765 2.





```

RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 2-682 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92115567; PubMed=1766878;
RA Sadosky A.B., Gray J.A., Hill C.W.;
RT "The RhsD-E subfamily of Escherichia coli K-12.";
RL Nucleic Acids Res. 19:7177-7183(1991).
RN [4]
RP REVISION TO 442.
RA Hill C.W.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP REVIEW.
RX MEDLINE=95020608; PubMed=7934896;
RA Hill C.W., Sandt C.H., Vlazny D.A.;
RT "Rhs elements of Escherichia coli: a family of genetic composites
each encoding a large mosaic protein.";
RL Mol. Microbiol. 12:865-871(1994).
CC -!- FUNCTION: Rhs elements have a nonessential function. They may play
an important role in the natural ecology of the cell.
CC -!- SIMILARITY: BELONGS TO THE Rhs FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X60998; CAA43309.1; -
CC EMBL; AE000242; AAC74538.1; -
CC EMBL; D90785; BAA15087.1; -
CC EMBL; D90786; BAA15094.1; -
CC EMBL; D90787; BAA15106.1; -
CC EMBL; L19083; AAB47716.1; -
CC PIR; C64898; C64898
CC EcoGene; EG10010; rhsE.
CC InterPro; IPR001826; Rhs.
CC Pfam; PF03527; Rhs; 1.
CC Pfam; PF05593; Rhs repeat; 3.
CC PRINTS; PR00394; RhsPROTEIN.
CC Multigene family; Complete proteome.
CC SEQUENCE 682 AA; 77141 MW; P704868D6E15601B CRC64;

Query Match 2.1%; Score 188.5; DB 1; Length 682;
Best Local Similarity 20.9%; Pred. No. 0.0092;
Matches 147; Conservative 83; Mismatches 212; Indels 263; Gaps 37;

Qy 884 LSRLTQSDGGRVWAYTYSAGNDQCPSTVITPDQGFTHQYQPELDDAVLQVANSNIT-- 941
Db 5 LTLTSHSTSEGRV-----SVHYGY-----DD-----KGLTGE 32
Qy 942 QQFSYNPVTGALLKAVAGQSGLTPIYPSG---RLKMNENDMKMSY-LWTLRGLNGY 997
Db 33 RQTVENPETGELLWHETGHA---YNEQGLANRVTPDSLPPVWLTLYGSGYLAKMGLG 88
Qy 998 TDLTG-TIQKTSRDTHGRVTQI--KDSSIKTLNYDDL-----NRHIGSQVTLATGHMLT 1050
Db 89 TPLLEFTRDLRHRTVTSFGSMAGSNAAKLTSTYPAGQLQSOHLNSLYD----- 140
Qy 1051 TTVEFDGLNRIEIGRLCDSSGHTLIDIQSWLKTQQLANRIVKLVNGLQRTQEQSYDSNRN 1110

```

```

Db 141 -----RDYG-----WMDNGDL-----VRIISGRQ-TREYGYSATGR 170
Qy 1111 LNQYKCDGAE-----PTDKYGHST-----VTQN-----FTYDIYGN 1142
Db 171 LSVRTLAPDLDIRPYATDPAGNRLPDDELHPDSTLTWPDNRIAEADAHVYRDEYGR 230
Qy 1143 ITACHTTFADG---TEDHATFRFANFTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINIT 1199
Db 231 LTEKTDRIIPAGVIRTDERT-----HHYH-----YDSQHELVEFT 265
Qy 1200 D-NHGN---TENFTYDTLGR-----LQNGQSVYGYDPLNRLVSQK 1236
Db 266 RIQHGEPLVESRYLYDPLGRNRMAKVRERDLTGMSLSRKRPEVTWTGWDG-DRLTTVQ 324
Qy 1237 TDTLCELYYR-----ETMLVNEVRNG-----EMIRL 1263
Db 325 TDTTRIQTVEPGSFTPLIRVETENGEREKAQRSLAETLQOEGSENGHVVFPAELVRL 384
Qy 1264 LRTGETIIAQPAS-----KVLITGTSQ---QSVIL 1292
Db 385 LDRLEEIRADRVSSERAWLAQCGLTVEQLARQVEYTPARKVHFYHCDHGLPLALI 444
Qy 1293 TSDKQNLQSEAYSAVKEKSTANDASI---LGYNGERADPVSGVTHLGNVRSYDPTLMR 1349
Db 445 SEDGNTAKRGYDEWGNQNLNENPHLHPYRLPQOHDSESGLYY--NRHHYDPLQGR 502
Qy 1350 FHTPDSLPFG-AGGINPYSYCLGDPINRSPGSHLSQAWTGMGIAGLILLTIATGGM 1408
Db 503 YITPD---PIGLRGGMNMYQVPL-NPIQVIDPMG-----LDAIENNTSGGL 544
Qy 1409 ATAAAGGIAAATASSTT--ALAFGALSVTSDITSIVSGALESDAS 1451
Db 545 -IYAVSGVPLIANSITNSAYQFG-----YDMAIVGGANGAA 583

RESULT 15
ID YDBA ECOLI STANDARD; PRT: 2003 AA
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydba.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12; PubMed=9097039;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]

```

RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RT "Multiple IS insertion sequences near the replication terminus in  
RT Escherichia coli K-12";  
RL Biochimie 73:1361-1374(1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS8 (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
CC BETWEEN AMINO ACIDS 839 AND 840.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AE000237; AAC74483.1; ALT SEQ.  
DR EMBL; AE000237; AAC74483.1; ALT SEQ.  
DR EMBL; D90778; BAA15009.1; ALT SEQ.  
DR EMBL; D90778; BAA18880.1; ALT SEQ.  
DR EMBL; D90779; BAA18881.1; ALT SEQ.  
DR EMBL; X62680; -; NOT ANNOTATED\_CDS.  
DR EcGene; EG11307; ydEA.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 489 I -> V (IN REF. 2).  
FT CONFLICT 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12CB853220EE CRC64;  
  
Query Match 2.1%; Score 186.5; DB 1; Length 2003;  
Best Local Similarity 18.3%; Pred. No. 0.065;  
Matches 251; Conservative 157; Mismatches 457; Indels 507; Gaps 59;  
  
Qy 468 PPAGEVDN--CPPPEYGFTRVKIIQTPVDSEPKDPKPIQVRVSLIGSQSHVILKIE 525  
Db 106 PPDESDDTPVPFPGG-----DEII--PDPDDTPPPKPVSNFNDVILDKTEKLTIR 158  
  
Qy 526 ERHYSATQLNLSL-----FOYNTDKSEL 549  
Db 159 DSVFTYFENADGTLSLOQSNRKAITNLWQIDEANNVVALEGVASDQATKWQYN--HNGEL 217  
  
Qy 550 GRLLKQTECKGKNGKTVSVVHKFTYK-----QDPTLOOS---HSI--TTHDNFTI 596  
Db 218 --VTGDNATVNNKGTIVDKDSTGTEINGNNGKVIQDGLDVSGGCHGIDITGDSATV 275  
  
Qy 597 HRSQVRSRYTRLFSDDTKDIVTQMSYDK--LGRLLTRILNSGTPYANTLTVDYELNNL 654  
Db 276 DNKGT-----MTVTDPESMGQIDGDKAIVNNEGESTITNGGTGTQINGDDATANN 327  
  
Qy 655 QDDNRPPPVITTTDVGNGQLRNEFDGAGRHSVQCLKDS-----GDGKFTYHTQOYDE 708  
Db 328 KTTVDGKDSGTGTEINGN-----GKVIQDGLDVSGGCHGIDITGDSATVDN 375  
  
Qy 709 QGRHHTSYSYLNGRQQTDPKVLHSMKS---YDNWQAIANTHSYGV----- 756  
Db 376 KG---TMTV-----TDPESIGIQVDGQAVNNEGESAITNGGTGTQINGDDATA 422  
  
Qy 757 --SEKITVDPIITLTKQLQSNNNVQTKGVTTVTPSQPIQITLFDAGHLQSCHTLT 814  
Db 423 NNGKTTVDGKDSGTGTEIAGNNGKVIQD----- 451  
  
Qy 815 RDGMDRVRKETDAIQCTIYQYDNNVNRVIOITLPGTIIVNRKYAPFSTDLITIRNGI 874  
Db 452 -----DLDVSG-----GHGIDITGDSATVDNK-----GTMT--VTPDESIGI 487  
  
Qy 875 SL-GQTF---DGLSRLTQSQDGRVWAYTYTSAGNDQCPSTVITPDQGFHYQVQPELDD 930  
Db 488 QIDGQAIIVNNEGESTITNGGTGTQINGDDATANN-----SGKTTVDGK-----DS 533  
  
Qy 931 AVLQVASNEITQQFSYNPVTGALLKAVAEQGSQSLPIYYPFSGRLKVENINDMKMSYMLWL 990

Db 534 TGTKIAGN-----IGIVNLDGSLT---VTGGAHGVENIGD----- 565  
Qy 991 RGLNGYVDTLTGTTQKISRDRTHGRVQTQKDSIIKITLNYDDLNRHISQVTDLATGHMLT 1050  
Db 566 ---NGTVNNKGD1--VVSdT--GSIGVLINGEGATVNTGDN-----VSNEATGFSIT 612  
Qy 1051 TT-----VEFDGLN-----REIGRKL-----CDSSGHTLDTIQ 1078  
Db 613 TNSGKVSLAGSMQVGDFTGVDLNGNNNSVTLAAKDLKVVQKATGINVSGDANTVNI-- 670  
Qy 1079 SWLKTQQLANRIVKLNGVLQRTQYSYDSNRNLNNOYKCDGAECECTDKYGHISIVQNFYD 1138  
Db 671 -----TGNVLVDKDKTADNAAEYFFDFSVGINVYGS-----NNVTL 708  
Qy 1139 IYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPNIRLYKDKAGRVINI 1198  
Db 709 --GKUTVV-----SDSEVTSRQSLPFG-----SAAKTSGLVVI 740  
Qy 1199 TDNHGNTENFTYDTLGRQLONGQGSVGYVDPLNRLVSOQKTDLDCELYYRETMVNEVRNG 1258  
Db 741 GD--GNTVMN-----GGLE-----LIGEK-----NALADG 764  
Qy 1259 EMIRLLRTGETIIAQRASKVLLTGTDSQOSVILTSQKQLSQRAYSAYGKHKSANDAS 1318  
Db 765 SQVTSIRTY-----SYTSVIVVSGES-----SVYLANGDTTISGEF 800  
Qy 1319 ILGNGERADPVSGVTHLNGVRSYDPTLMRFHTPDSLSPFAGGIN-----PYSYCLGDP 1374  
Db 801 PLGFAGVIRVQDKALLEISG-----ATLMQDIDSFEHGHTRTVEIQNLGFAFVTGEN 854  
Qy 1375 INRSPSGHLSWQAWTIGMGIGALL-----LTATGGVAIAAAGGIAAIA-- 1421  
Db 855 T-----TGINSGTISLLQNGKDPAPSPVILLATNGGSAATNAGTITGKVTQ 900  
Qy 1422 -----STSTALAFGALSVTSDTISVSGALEDASPKASSILGWSM--GMGAAGL--- 1470  
Db 901 HSFNKYSTGTS-----NSFIFNDVSSITGLVAQSNSTIINTDSGIIDLGRGSGVGLAI 956  
Qy 1471 -----ABSAIKGGTKLATHLGAFAEDGENA----- 1495  
Db 957 ADSTAENQKITLDSMWVDANDTTAMRDIASNAIDFGTGVGVGSDSYSGAGKNATAINQ 1016  
Qy 1496 -----LLKSTSESRIKMGVTRSLDRIVRNE-----GQVIKDHSGRYTD 1536  
Db 1017 LGGVITINAGAGMAAYGASNTVINCGTINLEKNGYDDSLAANTLVGNVAVYEHGTAIND 1076  
Qy 1537 N-----FMGKGEQAILVHGDKDGFLYHTEGKNHNGKGPYTRHTPEQLVDYLK 1583  
Db 1077 QTGVININVGTCQAFVNDGTGTIVNYGTICTFQVCSQSGNEYNTDDFTS-----LIYTG 1130  
Qy 1584 DNNIVDLTQGDKPVHLLSCYKSSGAAQKMAKYNRPVIAYSNKPITISQGL 1635  
Db 1131 GPTI---TRSGETV-----LNKSAVTDKLAGNVNNSGTLSDGQITVSSGL 1174

Search completed: July 3, 2004, 05:45:28  
Job time : 53 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2004, 05:45:33 ; Search time 68 Seconds  
(without alignments)  
7762.676 Million cell updates/sec

Title: US-09-889-874A-23

Perfect score: 8879  
Sequence: 1 VIKELKLFRRITMSDNEF.....PKIILGTEKTVKPTKTRP 1673

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8815	99.3	1660	Q9EVR7	Q9EVR7 xenorhabdus
2	1614	18.2	1562	Q883V9	Q883V9 pseudomonas
3	1553	17.5	773	Q8S6T1	Q8S6T1 coxiella bu
4	1530.5	17.2	774	O52880	O52880 coxiella bu
5	1416	15.9	1632	Q881P7	Q881P7 pseudomonas
6	1214	13.7	709	Q45948	Q45948 coxiella bu
7	1037	11.7	1290	Q881P2	Q881P2 pseudomonas
8	932.5	10.5	982	Q883W6	Q883W6 pseudomonas
9	920.5	10.4	1669	Q87VG6	Q87VG6 pseudomonas
10	919.5	10.4	505	Q45905	Q45905 coxiella bu
11	919.5	10.4	526	Q9X826	Q9X826 coxiella bu
12	916.5	10.3	526	O52883	O52883 coxiella bu
13	912.5	10.3	528	Q45949	Q45949 coxiella bu
14	884	10.0	589	O45902	O45902 coxiella bu
15	648	7.3	762	Q93JF8	Q93JF8 vibrio sp.
16	478.5	5.4	310	Q45904	Q45904 coxiella bu

17	450.5	5.1	2217	17	Q8TP72
18	429.5	4.8	336	2	Q45946
19	420.5	4.7	2364	16	Q82R58
20	395	4.4	820	16	Q82QR1
21	393.5	4.4	1976	16	Q8D4R5
22	392	4.4	1250	16	Q82RX1
23	391.5	4.4	2370	16	Q82RE3
24	388.5	4.4	1826	16	Q987Z7
25	386	4.3	2167	16	Q92EK5
26	386	4.3	2224	16	Q81U00
27	382.5	4.3	1117	16	Q8CK70
28	381	4.3	2183	16	Q86585
29	376.5	4.2	2082	16	Q9ACP4
30	365	4.1	1400	16	Q8X2F8
31	362	4.1	1411	2	Q46748
32	361	4.1	1510	16	Q8PHK6
33	359	4.0	380	16	Q87UI5
34	354	4.0	1710	16	Q8XTG0
35	353.5	4.0	332	16	Q88AD8
36	353.5	4.0	843	16	Q82950
37	352.5	4.0	1410	2	O52673
38	348.5	3.9	1394	2	O52668
39	348	3.9	1381	16	Q87PI5
40	341	3.8	1394	16	Q8X385
41	340	3.8	1316	16	Q8A692
42	338	3.8	1426	2	Q93V17
43	338	3.8	2306	16	Q8E107
44	337.5	3.8	2321	16	Q8F5B9
45	336.5	3.8	1404	16	Q8XED9

## ALIGNMENTS

### RESULT 1

Q9EVR7 PRELIMINARY; PRT; 1660 AA.

AC Q9EVR7; (TREMREL. 16, Created)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)  
DT 01-OCT-2003 (TREMREL. 25, Last annotation update)  
DE Nematocidal protein 2.  
GN XNP2.  
OS Xenorhabdus bovienii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Xenorhabdus.  
OX NCBI\_TaxID=40576;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=I73;  
RA Morgan J.A.W., Ousley M., Ellis D., Jarrett P.;  
RT "Novel toxins from Xenorhabdus strains with activity against nematodes";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ296653; CAC19493.1; --  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RHS\_repeat; 5.  
DR TIGRFAMs; TIGR01543; YD\_repeat\_2x; 12.  
SQ SEQUENCE 1660 AA; 185738 MW; AFABEA20AD70B164 CRC64;

Query Match 99.3%; Score 8815; DB 2; Length 1660;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 MSDNNEFTQANNFTSAVSGGVDPRGTGLYNTQITLGHIVGNGNLGPTPLTLTSLVSPLNKT 73  
DB 1 MSDNNEFTQANNFTSAVSGGVDPRGTGLYNTQITLGHIVGNGNLGPTPLTLTSLVSPLNKT 60

QY 74 DIGFGIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLQOKKLDNLRTEKDKENY 133  
DB 61 DIGFGIGNFGLSVYDRKNSLLSLSTGENYKVIETDKTKLQOKKLDNLRTEKDKENY 120

QY 134 RIHKSGDIEVLGTGFNNNAFDLKVPKLLNFAGHAIYIDWNFEATQPRNLRIYDLDGDH 193

Db 121 RIHKSGDIEVLGTGNNNAFDLVKPKKLLAPAGHAIYIDWNFEATQRLNRIYDDLDGHD 180  
Qy 194 IPLNLEYQGLIKITLTLFPQKEGYRTELRFNLQNSIHNFSLGNENPLTWSFGYTP 253  
Db 181 IPLNLEYQGLIKITLTLFPQKEGYRTELRFNLQNSIHNFSLGNENPLTWSFGYTP 240  
Qy 254 GKNGILGOWITSMAPGLKETVNSNNQCHHPQSANLPVLYVTLKMQVQACOPAI 313  
Db 241 GKNGILGOWITSMAPGLKETVNSNNQCHHPQSANLPVLYVTLKMQVQACOPAI 300  
Qy 314 QAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNGSTESRRYKQKEGHDQIVRIERY 373  
Db 301 QAEYSYTSNHYVGGSGNGIWNKLDNLYGLMTEYNGSTESRRYKQKEGHDQIVRIERY 360  
Qy 374 NYHLLTSECKQNGYIOTTETAYAIIGHNFDSPQSFQPKTKTETWRSADNSRSEI 433  
Db 361 NYHLLTSECKQNGYIOTTETAYAIIGHNFDSPQSFQPKTKTETWRSADNSRSEI 420  
Qy 434 TETTFDSGNPLTKVKKDKTKTKIISPSHWEYYPAGEVDNCPPEYGFTRVKKIIOT 493  
Db 421 TETTFDSGNPLTKVKKDKTKTKIISPSHWEYYPAGEVDNCPPEYGFTRVKKIIOT 480  
Qy 494 PYDSEFDDPEKFTQYRYSIGQSHVTLKIEBHYSATOLLNSTLFOYNTDKSELGRLL 553  
Db 481 PYDSEFDDPEKFTQYRYSIGQSHVTLKIEBHYSATOLLNSTLFOYNTDKSELGRLL 540  
Qy 554 KQECTKGNGKTVSVVHKFTYTKQDDTLQOSSHITTHDNFTIHRQVRSRYTGRFSDT 613  
Db 541 KQECTKGNGKTVSVVHKFTYTKQDDTLQOSSHITTHDNFTIHRQVRSRYTGRFSDT 600  
Qy 614 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLTVDYELANLQDDNRPFPVITTDVNGNQ 673  
Db 601 DTKDIVTQMSYDKLGRLLTTLNSGTPYANTLTVDYELANLQDDNRPFPVITTDVNGNQ 660  
Qy 674 LRNEFDGAGRHVSQCLKSDGCKEFTIHTQQVDEGRHHTSYSDYLTNGRQCTDPDKV 733  
Db 661 LRNEFDGAGRHVSQCLKSDGCKEFTIHTQQVDEGRHHTSYSDYLTNGRQCTDPDKV 720  
Qy 734 HLSMSKSYDNWQGIANTHWSYGVSEKITVDPIITLTKQLQSNNNVQTKVTTVTPSQ 793  
Db 721 HLSMSKSYDNWQGIANTHWSYGVSEKITVDPIITLTKQLQSNNNVQTKVTTVTPSQ 780  
Qy 794 OPTQITLFDAGHLSQCHTLTRGWRVRKETDAIGCCTIYQVDNVRVITLDPGTTIV 853  
Db 781 OPTQITLFDAGHLSQCHTLTRGWRVRKETDAIGCCTIYQVDNVRVITLDPGTTIV 840  
Qy 854 NRKYAPSTDTLITDIRVNGISLQQTQDGLSRLTQSGGRVWATYSAGNDQCPSTVI 913  
Db 841 NRKYAPSTDTLITDIRVNGISLQQTQDGLSRLTQSGGRVWATYSAGNDQCPSTVI 900  
Qy 914 TPDQGFTHYQBELDAVLQVANSNEITQOFSNPVTGALLKAVAGQSLTIPIYPSGR 973  
Db 901 TPDQGFTHYQBELDAVLQVANSNEITQOFSNPVTGALLKAVAGQSLTIPIYPSGR 960  
Qy 974 KMENINDMKMSYLWTLRGLNGYTDLTGTIKISRDTHGRVTOIKDSSIKTLLNVDL 1033  
Db 961 KMENINDMKMSYLWTLRGLNGYTDLTGTIKISRDTHGRVTOIKDSSIKTLLNVDL 1020  
Qy 1034 RHIGSQVTDIATGMLTTTVEPGLNREIGRKLCDSSGHTLDDIQSWLTKTQOLANRIVKL 1093  
Db 1021 RHIGSQVTDIATGMLTTTVEPGLNREIGRKLCDSSGHTLDDIQSWLTKTQOLANRIVKL 1080  
Qy 1094 NGVLQRTQESYSDSRNLNQYKCDGACPTDKYGHISIVTQNFYDIYGNITACHTTFADG 1153  
Db 1081 NGVLQRTQESYSDSRNLNQYKCDGACPTDKYGHISIVTQNFYDIYGNITACHTTFADG 1140  
Qy 1154 TEDHATFKFANPTDPCQLTEVHHTPMDPNILKYDKAGRVINITDNHGNTEFTYDTL 1213  
Db 1141 TEDHATFKFANPTDPCQLTEVHHTPMDPNILKYDKAGRVINITDNHGNTEFTYDTL 1200  
Qy 1214 GRLONGGSSVGYDPLNRLVSQCTDLDCELYYRETMVNEVRNGEMIRLLTGTETIAQ 1273

Db 1201 GRLONGGSSVGYDPLNRLVSQCTDLDCELYYRETMVNEVRNGEMIRLLTGTETIAQ 1260  
Qy 1274 QRASKVLLTGTDSQOSVILTSDKQNLQOEAYSAYGKHKSTANDASILGYNGERADPVSGV 1333  
Db 1261 QRASKVLLTGTDSQOSVILTSDKQNLQOEAYSAYGKHKSTANDASILGYNGERADPVSGV 1320  
Qy 1334 THLNGVSYDPTLMREHPTDLSLPPGAGGINPVSICLGDPIKNSDPSGHLSSQAWTIG 1393  
Db 1321 THLNGVSYDPTLMREHPTDLSLPPGAGGINPVSICLGDPIKNSDPSGHLSSQAWTIG 1380  
Qy 1394 MGIAGLLTIATGGMAIAAAGGIAAAIAASTTTALAFGALSVTSDITSIVSGALEDASP 1453  
Db 1381 MGIAGLLTIATGGMAIAAAGGIAAAIAASTTTALAFGALSVTSDITSIVSGALEDASP 1440  
Qy 1454 ASSILGWVSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSSRIKMGVTRS 1513  
Db 1441 ASSILGWVSMGGAAGLAESAIGKGTKLATHLGAFAEDGENALLKSTSSRIKMGVTRS 1500  
Qy 1514 LDREIVRNEEGQVTKDHSRGYTDNFMKGGEQAILVHGDGKDFLYHTEGNKHNGKGPYTRH 1573  
Db 1501 LDREIVRNEEGQVTKDHSRGYTDNFMKGGEQAILVHGDGKDFLYHTEGNKHNGKGPYTRH 1560  
Qy 1574 TPEQLVDYKDNNTVDLTQGGDKPVHLLSCYKSGSGAADMVAKYINRPVTAYSNKPTISQ 1633  
Db 1561 TPEQLVDYKDNNTVDLTQGGDKPVHLLSCYKSGSGAADMVAKYINRPVTAYSNKPTISQ 1620  
Qy 1634 GLARIERKDFLTKSTYHSYDPRKIILGRTEKTVKPKTRP 1673  
Db 1621 GLARIERKDFLTKSTYHSYDPRKIILGRTEKTVKPKTRP 1660

RESULT 2  
Q883V9  
ID Q883V9 PRELIMINARY; PRT; 1562 AA.  
AC Q883V9  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE YD repeat protein.  
GN PSPT02239.  
OS Pseudomonas syringae (pv. tomatop).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkac L., Bearan M., Haft D., Selengut J., Neilson W., Davidson T.,  
RA White O., Fraser C., Collier A.;  
RT "Complete sequence of Pseudomonas syringae."  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016863; AA055755.1;  
DR TIGR; PSPT02239;  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0003910; F:DNA ligase (ATP) activity; IEA.  
DR GO; GO:0006310; F:DNA recombination; IEA.  
DR GO; GO:0006281; P:DNA repair; IEA.  
DR GO; GO:0006260; P:DNA replication; IEA.  
DR InterPro; IPR000977; DNA\_ligase.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RBS repeat; 6.  
DR PROSITE; PS00697; DNA\_LIGASE\_A1; 1.  
KW Complete proteome.  
SQ SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BFE37BF1 CRC64;  
Query Match 18.2%; Score 1614; DB 16; Length 1562;  
Best Local Similarity 30.3%; Pred. No. 1.4e-78;  
Matches 490; Conservative 239; Mismatches 700; Indels 190; Gaps 51;  
Qy 14 MSDNNEFFTOANNFTSAVSGVDPRGLYNIQTILGHIVGNLGLPTLPLTSLYSPLNKT 73

Db	1	MTTSTSVHNAFNMSYLSQSGVDPRGTQGVTSINLPEVKSNGRLGPPVPLVLSYNPLNVQ	60
Qy	74	DTGPGTGFNPGSVVDKXNLSLSGENYKVI--ETDTKVLQOKKLNLRPEKDLKENC	132
Db	61	DSGFLGWNLQSQSYDPGTRIVSLSGGETFKVGEISGLDQLMPEKKLSFHYKO--DTR	119
Qy	133	YRIHKSGLDIEVLTPGNNAFDLVPKLLNPAGHAIYIDWNFEATQPLANRIYDDLGH	192
Db	120	YRWVHKSQGVEEVLVSLGNRIALPVRISPEGHGITHUYASFQAYQMLSEVWD--DGQ	178
Qy	193	DIPLLNLEVOGLIKTILTFPQKEGYRTELAPLARQLNSIHNFSLGNENPITWISFGYTP	252
Db	179	VILITIRD----STSVRLLYGAPKADAEFVMLGSENENVARIELPTANKASWRFTYSI	234
Qy	253	IGKNILGOWITSMTAPGGLKETVYNSNNQGHFPQSANLEPVLPIVYTLMKQVPGAGOPA	312
Db	235	IRGHSC----IASVDTPVGHEDEVFYQDS--GHQFPLSAGREPLPRVTRHLTTPGFLQPE	288
Qy	313	IQAEYSY----TSHNYVGGNGNIW--NNKLDNLICLM--TNYNGVSTESRRYKKEGHD	364
Db	289	VDVRYAYKDAGRERNFLGAGLDIAWEDNGLDNLRYLGAAPLYLSSTETLWVNDV--	345
Qy	365	QIVRIERTYNNVHLLTSECKQONGYIOTTETAYALIGHNFDQSQSQFOLPKTKTETWR-	423
Db	346	-VRSIERVFNQFHLLALETRQNLSTLEVDTRYIIEEGKFPDLOPNYCOLPREVATTWEL	404
Qy	424	SADNSY-RSEITETTFDEGNPLTKVIOKTKQKILSPSTHWEYPPAGEVNCPEPYG	482
Db	405	SPDGSVPEIETVSDSYDSYCNLLAQTQANGVIETSS-----EWISVSGE-DGCPDPDG	456
Qy	483	FTFEPVKIKITQPYDSSEFKDDPEKFIQVRY----SLIGSOSHVLTKIE-----ERHVSATQ	533
Db	457	FVETLKAKSVVPAQSDYGHALVULTRYRYKALPALAGSGONLWAAESETLQQTTDGEK	516
Qy	534	LLNSTLIFYNTDKS----ELGRLLKTECTKGENGKT---YSVHHKFTYTKODTLQOSH	586
Db	517	ELQQTITYIEDNPYDAFOYGRIFHQSVTLEGLSTTTDVRYLLOQDPD---DQIVQQVTV	573
Qy	587	SITHDNFTIHRQVR---SRYTGRIFSDDTKDIIVTQMSYDKLGRLLTRTLN--SCTPYA	642
Db	574	QIVT--GFDMQTKVILEHSLFTGEPLLRDNDVEIRYDYNLRVAVSETVSPNKEEYK	631
Qy	643	NLTUDYELNLOODNRPPVITTTDVNGNLKNEPFDGAGRHSQCCLKSDGSGKFYTIH	702
Db	632	ATHRYEVLCAVTKDQAEQLF---DVKNQVITSRFDGLGRVIYEARADADNPVHRLD	688
Qy	703	TQQ-----YDEQGRHHTSVSYLTNGRQOTDDPKVHLSMSKSYDNWQIANTHWSYGV	757
Db	689	LRQTYEAAYPDAGDKVBEETSYDML-----DOCKEALINYFYDDMQQLSVTGPDVT	741
Qy	758	EKITVDPITLAT-----KOLQNSNNVOTGKEVTY--TPSQOPIQITLFDAGHLQSCH	811
Db	742	TIGQTDVGTQAGNPLOQRWTESNDGLQTSVSETEWNLNLFDEPTSVRLDRDLMLSEPV	801
Qy	812	TLTR--DGWDRVRKETDAI---GQCTIYOYDNNRVIQITLPDGTIVNRKYAPFSTDIL	865
Db	802	SLSRYQDGLGLRVKVESGLPIRERSTTYGYDVFDRVTANTLPDGAVAVRRVAPHSGEDL	861
Qy	866	ITIRV--NGIS--LGOOTFDGLSRLTQSDQGRVWAYTYVSAGNDCQPSVTITPDQGFTH	921
Db	862	PAMIGVDHNGKSVLGEQKFDGLDRIVTSTITGRRERELSYISDLMQ--PKTVKLPSGRQID	920
Qy	922	YQVQPELDDAVLQ-----VASNEBITQOFSYNPNVTGALLKAAVAGOSLTPYIYPSGR	975
Db	921	YDYLPELGDPELKETQSDTVARUKLTADYIDVQNAFLIGSSBQGEELQREYVYSTGLKS	980
Qy	976	ENINDM---KMSVILWTIRGLENGYTDLTGTIQKISRDTGRTVQIKDSSIKITLNYDDL	1032
Db	981	EQRTSQGIENTHMYRYSRGLGLP-SYIDVLVQEQQLSVYDDFGRLAQTSLGVEVSDFTYDTF	1040
Qy	1033	NRHIGSQVTDLATGHMLTMTTTFEPDGLNREIKRCLDCSSGHTLTDIQOSWLKTOQLANRIVK	1092

Db 1041 GRTASTIATDSSNGGVVLSLEYDAQREAOQRTF--TINGANQOMQVQVDDVDQMKVKTLS 1099  
 Qy 1093 LNCVLQTEQYGYDSRNRNLNOYKCDGAECTDKYGHISIVTQNTFYDIGNITFACHTTTAD 1152  
 Db 1100 EGAVIIFREHYGYDLGRLTYDCSGKQRPVDPYGMTISRQVFSFDGLNNLTLVTTF--D 1158  
 Qy 1153 GTEDHATFPKANPTDPCOLTEVHHH-----PMDPNIRLKYDKAGRVIN 1197  
 Db 1159 GGNRRARY-FYEGIDPAQUTRVNTTOLAWNARLLPVKNKNDSTYPEIRLTYDPDGNL-- 1215  
 Qy 1198 ITDNHGNWNTFTYDTLIGRL---QNGOGSV-YGYDPLNRLVRSOKTDTLDCELYRETMLVN 1253  
 Db 1216 ITDEADRL--LSYDPLGRLLVSPAGDVRYRYPQDRLAGETG---QRFVRDGVLAS 1269  
 Qy 1254 EVRNGEMIRLLRTGETIIIAQORASKVLLTGTSSQSVILTSQKNLSQEAYSAYGKHKT 1313  
 Db 1270 QLGASQNSTVMRGDGYLLAEQQGSDALLFSTISNSVLSEVHPDGVNKRSTVYGHSSGD 1329  
 Qy 1314 ANDASITLNGYGERADPVSGVTHLNGYRSYDPTLMRFHTPDSLSFPGAGGINPYSYCLGD 1373  
 Db 1330 DPFAGRLGYNGELHETDTGWQLLNGYRAYNPVLMRFHSPDSWSPFEGGLNAYAYGED 1389  
 Qy 1374 PINRSDPSGH-----LSQAWTG-----IGMGIAGLLLTATGG 1407  
 Db 1390 SVANGVPTGHWGFKLPRIFFTRLEKQAWKADKKEFWRLIEODIKNEGLEGYAOAYRDL 1449  
 Qy 1408 MATAAAGG---IAAAATSTTALAFGALSVTSDITSIVSGALEDAS--PKASSILGWVS 1462  
 Db 1450 QAKSAKONKAYISLKVHREKALE--AETLKANTGDVMSRSRESIITPKAASSGSIS 1507  
 Qy 1463 MCMGAAGLAESAIAKGGTKLATHGAFAPDGENALLKSTSESSIRKKGVTRELSDREIVRN 1521  
 Db 1508 -----NLGGSFGSDRDVGLRNVSSA---KSEVLRQD--FVRN 1540  
 RESULT 3  
 Q9S671 PRELIMINARY; PRT; 773 AA.  
 AC Q9S6J1  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Coccidiella burnetii.  
 OG Plasmid QpDV.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 OC Coccidiellaceae; Coccidiella.  
 OX NCBI\_Taxid=777;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R31140;  
 RA Radomski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;  
 RT "Sequence of QpDV plasmid."  
 RL Submitted (PEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF131076; AAD33495.1;  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR008530; YD.  
 DR Pfam; PF05593; RKS\_repeat; 6.  
 DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 8.  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 773 AA; 83759 MW; 4C5FCE0481CCAC44 CRC64;  
 Query Match 17.5%; Score 1553; DB 2; Length 773;  
 Best Local Similarity 44.6%; Pred. No. 9.6e-76;  
 Matches 342; Conservative 119; Mismatches 275; Indels 30; Gaps 17;  
 Qy 741 YDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNNNVQTGKVITYTTPSQPIQITL 800  
 Db 4 YDSNGQNLTVFSDGYQERSYVDPIETRAAT--LQPSGSKQLGQLTETYNLAGLPIKVTQ 61  
 Qy 801 FDEAGHLQSCHTLTRGDWRVRKETDAIGCTIYQYDYNARVITQITLDCGTIVNKKYAPF 860  
 Db 62 YDSQGTQGSAAHYEDGLGRLKEDTDLGQVITYEYDFGRVQTTLPENTIQKSYAPH 121

```

QY 861 STDTLITDIRVNGISLGOOTFDGLSRLTQSODGGRVWAYTYSAGNDCPCSTVITPDQCFI 920
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 STASLITGISVNNFSMGNQTFDSLERLTETTSRGTSRTSAGTSAFSEYENAS-SVPAAVTAPTGETV 180

QY 921 HYQVPELDDAVLOVASNEITQOFSYNPVTGALLKAV-AGEQSLTPIIYPSGRLKME-NI 978
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SYEYLKELGNVAKKISAPEILKTWDYDALTGAMTSATQAAGMIRQMTYIYPSGLLKNETSM 240

QY 979 ND---MKKMSYLTWLRGLENGYDITGCTIOKISRDTHGRVTOIKDSSIKTTLNYDDLNR 1035
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PDGAAQKSTATYISLAGAPQSYTDVFGVTOYDYDEHGRIGIEDNDIKVSLGYDAFGRF 300

QY 1036 IGSQVDTLATGMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSWLKATQOLANRIVKNG 1095
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 TKQATDKTKTGAVLSTLTLYDDLNRKREIKREISASGOSVLVECTYQRNHLLKERITQGR 360

QY 1096 VLQRTQOYSVDSNRNLNOKYKDCGAECPDKYGHISIVTQNTFYDIYGNITACHTTFADGTE 1155
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 TTLKREVPAYDSNRNLLEYTCNGEARPDQPYGKAIRHQTFSDYDALGNMTKTQDPSGG-R 419

QY 1156 DHATFKFANPTDPCQTEVHHTHPDMPNIRLKYDKAGRVINITDNHNTENFTYDTLGR 1215
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 NTATYIY-SAIDPTQLLKVNNDHSDYKPEITLEYDKAGRMIR--DEAGRT--LRYDALGR 474

QY 1216 LQ--NG---QGSVGYDPLNRLVSKQT-DTLDCELYYRETMVNEVR--NGEMIRLLRTG 1267
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 LQOVNGAGAGKGGQYADALNTLVSVQVQDEPIYDLVYRADDLVGEARRDGSSTRYVKN 534

QY 1268 ETIIAQ---ORASKV-LLTGTDSQOSVILTSKQNLSEEA-YSAYGKHKSTANDASILGY 1322
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 GCCVQCTKQGSNTSLTITNQQGSVLSVSEGNHAPQDCIYTRYGYRTPQETPSVLGF 594

QY 1323 NGERADPVSGVTHLNGYRSDPTLMRFHTPDSLSPFGAGGINPYSCYCLGDPINRSDPSG 1382
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 NGERLDPVSGVTHLNGYRAYNPILMRFNCPDPSWSPFGAGGINPYAYCDGDPINRVDPNG 654

QY 1383 HLSQWQMTGIGMGTAGLLLTATGGMIAAAGGIAAIAASTSTTALAFAGLSVTSITS 1442
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 HLSQWQMTGIGMGTAGLLLTATGGMIAAAGGIAAIAASTSTTALAFAGLSVTSITS 1442

QY 1443 VSGALEDASPKASSILGWSMGMAAGLAE---SAIKGGTKLATHL 1485
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 ASGALEDANPQASATLGLWISLGLGPGAVSGLATARAAGKLLISGL 760

```

## RESULT 4

```

OS2880 PRELIMINARY; PRT; 774 AA.
AC OS2880;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN ORF 774.
OS Coccidia burnetii.
OG Plasmid QpRS.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coccidiaceae; Coccidia.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Priscilla Q177;
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15898; CAA75841.1;
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs repeat; 6.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 8.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 774 AA; 83774 MW; 918A5433D7991BD8 CRC64;

```

```

Query Match 17.2%; Score 1530.5; DB 2; Length 774;
Best Local Similarity 44.5%; Pred. No. 1.6e-74;
Matches 341; Conservative 120; Mismatches 275; Indels 31; Gaps 18;

QY 741 YDNWQIANTHWSYGVSEKITVDPIITLTKQLQSNVNTQKREVITYTSPQPIQITL 800
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 YDSWGNHLTVPSDCYQERSVVDPIRRTAT--LQESGSKLQOQLTEYNLAGLPIKVQ 61

QY 801 FDEA-CHLOSCHTLTFEDGWDVRKETAIGACTIYQDYNVNEVCIITLPCDCTVNRKYP 859
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 YNSQGTGGSAAHYDGLGQLRKETDELQGVTFEYDFHGRVITQTLTPENTIIQKSYAP 121

QY 860 FSTDLITDIRVNGISLGOOTFDGLSRLTQSODGGRVWAYTYSAGNDCPCSTVITPDQCF 919
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 HSTASLITGISVNNFSMGNQTFDSLERLTETTSRGTSRTSAGTSAFSEYENAS-SVPAAVTAPTGET 180

QY 920 HYQVPELDDAVLOVASNEITQOFSYNPVTGALLKAV-AGEQSLTPIIYPSGRLKME-N 977
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SYEYLKELGNVAKKISAPEILKTWDYDALTGAMTSATQAAGMIRQMTYIYPSGLLKNETS 240

QY 978 IND---MKKMSYLTWLRGLENGYDITGCTIOKISRDTHGRVTOIKDSSIKTTLNYDDLNR 1034
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 MEDGAAQKSTATYISLAGAPQSYTDVFGVTOYDYDEHGRIGIEDNDIKVSLGYDAFGR 300

QY 1035 HIGSQVDTLATGMLTTTVEFDGLNREIGRKLCDSSGHTLDIQOSWLKATQOLANRIVKLN 1094
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 FTKQATDKTKTGAVLSTLTLYDDLNRKREIKREISASGOSVLVECTYQRNHLLKERITQGR 360

QY 1095 GVLRTEQYSVDSNRNLNOKYKDCGAECPDKYGHISIVTQNTFYDIYGNITACHTTFADGT 1154
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RTTLKREVPAYDSNRNLLEYTCNGEARPDQPYGKAIRHQTFSDYDALGNMTKTQDPSGG- 419

QY 1155 EDHATFKFANPTDPCQTEVHHTHPDMPNIRLKYDKAGRVINITDNHNTENFTYDTLGL 1214
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 NTATYIY-SAIDPTQLLKVNNDHSDYKPEITLEYDKAGRMIR--DEAGRT--LRYDALG 474

QY 1215 RLO--NG---QGSVGYDPLNRLVSKQT-DTLDCELYYRETMVNEVR--NGEMIRLLRT 1266
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 475 LQOVNGAGAGKGGQYADALNTLVSVQVQDEPIYDLVYRADDLVGEARRDGSSTRYVKS 534

QY 1267 GETIIAQ---ORASKV-LLTGTDSQOSVILTSKQNLSEEA-YSAYGKHKSTANDASILG 1321
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 535 NGCCVQCTKQGSNTSLTITNQQGSVLSVSEGNHAPQDCIYTRYGYRTPQETPSVLG 594

QY 1322 YNGERADPVSGVTHLNGYRSDPTLMRFHTPDSLSPFGAGGINPYSCYCLGDPINRSDPS 1381
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 595 NGERLDPVSGVTHLNGYRAYNPILMRFNCPDPSWSPFGAGGINPYAYCDGDPINRVDPN 654

QY 1382 HLSQWQMTGIGMGTAGLLLTATGGMIAAAGGIAAIAASTSTTALAFAGLSVTSITS 1441
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 HLSQWQMTGIGMGTAGLLLTATGGMIAAAGGIAAIAASTSTTALAFAGLSVTSITS 1441

QY 1442 IVSGALEDASPKASSILGWSMGMAAGLAE---SAIKGGTKLATHL 1485
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 IASGALEDANPQASATLGLWISLGLGPGAVSGLATARAAGKLLISGL 761

```

## RESULT 5

```

OS2880 PRELIMINARY; PRT; 1632 AA.
AC OS2880;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN PP1882.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;

```





OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
OX Cocciliaceae; Cocciliella.  
CN NCBI\_TaxID=777;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=Nine Mile phase I;  
RT Thiele D., Willems H., Haas M., Krauss H.;  
RA "Analysis of the entire nucleotide sequence of the cryptic plasmid  
RT Ophi from Cocciliella burnetii";  
RL Eur. J. Epidemiol. 10:413-420(1995).  
RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=Nine Mile phase I;  
RA Thiele D.,  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.

DR EMBL; X75356; CAA53129.1; -;  
DR PIR; S38241; S38241.  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.

DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RHS\_repeat; 8.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 8.  
KW Plasmid.

SQ SEQUENCE 709 AA; 79417 MW; A377B9A2A72CCF86 CRC64;  
Query Match 13.7%; Score 1214; DB 2; Length 709;  
Best Local Similarity 40.8%; Pred. No. 2.1e-57;  
Matches 287; Conservative 101; Mismatches 280; Indels 36; Gaps 16;

QY 554 KQTECTGKNGKTSVVKHFTYTKQDDTLOQSHSIITHTHNSVRSRYGRFSDT 613  
DB 3 RRAEVLTSKGGKYNQNTTFAISSQAEHLQKIDFTGGDGKTSISREQSYGRLLSST 62  
QY 614 DTKDVTQMSYDKLGRLLTTLN-SGTPYANTLYDYELNNLODNRPPVITTDVNGN 672  
DB 63 DELGNVTQYDEYDEGRLLTQTVNASSTYASTRYSLSTDARGKVTAIKTTVDPKGN 122  
QY 673 QLRNEFGAGRHVSQCLKSDG-----DGKFTYHTQCYDEQGRHHTSTVSDYLTNGROQ 727  
DB 123 QLRYYDGLGNLQERLKDAAVSQKTGYWYTHQQQYDALGRESKITQDVL-----R 177  
QY 728 TDPKVV-----LSMSK--SYDNMGQIANHWSYGVSEKIITVDPIITATKQLQSNVNO 781  
DB 178 LDSEVGHAGSVLSTKVMHYDSWQNLTVFSDGYQERSYDPIITRAT--LQPSGSOK 235  
QY 782 TGKEVNTYPSQOPIQTLTDEAGHLOSCHTLTRDGDWRVRKETDAIGQCTIYQDYNVR 841  
DB 236 LGQOLTENYLAGLPKIVQDSQTEGSAHYEYDGLGURKETDELQITLLEYDHFGR 295  
QY 842 VIQITLPGDTIVNRKAPFSTDLITDIRVNGISLQOQTFDGLSRLODQGRVWAYTY 901  
DB 296 VTQTLPTENTIIQSYAPHSASTSLITGISVNNQSMGNTQFDSLERLTTSGGRTSAFSY 355  
QY 902 SAGNDCPSTVITPDGFIHYQPELDVAQLVANSNEITQPSYNPVTCALLKAV-AG 960  
DB 356 ENAS-SVPAAVTAPGTGTSTVEYLKNGNNAVKISAPILLQTDWYDALGTAMTSATQAG 414  
QY 961 QSLTPIYPSGRKME-NIND---MKMSYLWTLRGLNGYTDLTGTIQKISRDTHGRVT 1016  
DB 415 MIRQMTYPSGLLKNSTMPDGAQKSTAYTSLAGAPQSYTDVFGVTQYDYDEHGRRI 474  
QY 1017 QIKDSSTKTYLNDLNRHIGSQVTLATGCHMLTTTVEPCLNREICRKLCDSSGHTLDI 1076  
DB 475 GIEDNDIKVSLDYDAFRFTKQATDKKCAVLSLTLYDNLNREIKREISAGQSVLVI 534  
QY 1077 QQSLKLTQQLANRIVKLVQLQTEQSYDSRNRLNQYKCDGACETDKYGHISVTQNF 1136  
DB 535 EQIVQRNHLKERITQSGRTTLAKEMFAYDSRNLLIETCNGEARQDPYGRKAIHQTS 594  
QY 1137 YDIYGNITACHTTFAQTEDEHATFKFANPTDPCQLTEVHTHTDPDMPDNLKYDKAGRVI 1196  
DB 595 YDALGNMTKTQDPSGG-RNTATYIY-SAIDPTQLLKVANDHSDYKPEITLEYDKGRMI 652  
QY 1197 NITDNGHNTENTFTYDLGRLOQ--NG---QGSVYGYDPLNRLVSO 1235

DB 653 R--DEAGRT--LRYDALGRLOQVNGAGAKGGQYAYDVLTNLVSQ 692

RESULT 7

Q88LP2

ID Q88LP2 PRELIMINARY; PRT; 1290 AA.

AC Q88LP2;

DT 01-JUN-2003 (TREMELrel. 24, Created)

DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Hypothetical protein.

GN PPI887.

OS Pseudomonas putida (strain KT2440).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=160488;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22423060; PubMed=12534463;

RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,

RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,

RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,

RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,

RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzar A.,

RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,

RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,

RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuenmler B.,

RA Fraser C.M.;

RT "Complete genome sequence and comparative analysis of the

RT metabolically versatile Pseudomonas putida KT2440.";

RL Environ. Microbiol. 4:799-808(2002).

DR EMBL; AEO16781; AAN67506.1; -;

DR TIGR; PPI887; -;

DR InterPro; IPR006530; YD.

DR Pfam; PF05593; RHS\_repeat; 6.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 1290 AA; 146578 MW; 7081A3F10B287F92 CRC64;

Query Match 11.7%; Score 1037; DB 16; Length 1290;

Best Local Similarity 27.7%; Pred. No. 2.2e-47;

Matches 375; Conservative 196; Mismatches 578; Indels 204; Gaps 52;

QY 14 MSDNNEFTQANNFTSAVSGGVDPRTGLYNIQITLGHVGNLGLPTLPLTSLSPNK 73

DB 2 MSTSTSVHNAFNSVIESGVDPRTQYTVSIRLPELQNDLQPGFELALFSP 61

QY 74 DIGFIGNFGSLVYDRKNSLLSTGENYKVI-ETDKTVKLOKLDNLRFKDLKEN- 131

DB 62 DSGFGKWN--LQTLQVRKHITVLSSETGITKTSVTGRLEMKQLKQF--DLYEDP 117

QY 132 -----CYRIHKSQDIEVLTFGNNAFDLVKPKLLNPAGHAIYIDW-NFEATQRLN 183

DB 118 PPGGAARFVREHRSGLVEVWVGSGEVALVELHSLPLGHLHLISYLPFGGHMRLS 177

QY 184 RIYDDLDGHDIPLNLNLYOGLIKTILTLFPQKEGYRTEIRF---LNRQLNSIHNFSLGN 240

DB 178 EVQDESD---VLLRLQSDNSRVELLYCPSGDGGDLARYAMTLESNDRVSEILPT 233

QY 241 ENPLTWSFGYTPICKNGILGQ-WITSMTPAGLKETVNSNNNGHFFPOSANLPV-LPY 298

DB 234 ANQARWFTY-----EDVLGYLCVRECTPYGGYVVFY--QDAGHKFPSSAARDKNLPR 286

QY 299 VTLMKVPGACQAPQAIQAEYSY-SHNVYGGSGNIGWNNK-LDNLVGLMTYNYGSTEESR 356

DB 287 VTRHEIDPRFQAAKVVRYEYPGTHNFIQGGSSISWSDGLDNLKYVPEDYTVKSIQI-- 344

QY 357 YKDEGHDOIVR-TERTYNNVHLITSECKQNGVIQTETAYVAIIGHNFDSPSQFQLP 415

DB 345 ---QEVRRQSVRTTRIFNRPHLLTEQATQCKLLOAFTRYADNAG-NFESSEPYQLP 400

QY 416 KTKTETWRSADN--SYRSEITETTFDESGNPLTKVTKDKTKQKIISPTHEWYYPAGEV 473

Db 401 HDTQKSLISNPSRQREKRVTRYDSHGVLTRLLPNQVLETVN-----WYGAABERG 453  
QY 474 DNPCEPPYGFTRFVK-KIIQ-----TPYSEFKDDPEKFIQYRY 511  
Db 454 DE-----HGFVRLKURTVQPAATGAGAAATLQAFQYRALTPGGS-----YLKQW 500  
QY 512 SLIGSQSHVTLKEERHYSATQILN--SLIFQ-YNTDKSELGRLKQTECTKG-ENGKTY 567  
Db 501 RLLESET-----LSEGSAPGTLEKISKLYQESAEBSYSGVRQOTVSYGVGDSGSPF 555  
QY 568 SVVHKFTYTKQDD--TLQOSHSTTTHDNFTIHSQVRSRYTGRFSDTDTKQIVTQMSYD 625  
Db 556 DTLTHYGSLPDDERALQTVKLVGVGSGSKTTLRLHALDTGEPLLNLDNGVEIRYYD 615  
QY 626 KLGRLTRTLNSGTPVANTLYDYELNLODDNRPFPVITTDVNGQLNRPFDGAGRV 585  
Db 616 ALARVTRVAVPKPFEAERNYKYLFCAYDNEQAQQWYDKQVETHTL---LDGLSRPI 572  
QY 686 SOCLKSDG--DGKFTYHTQOYDEGRHHTSYDYLTNGRQOTDPDKVHLSMSKSYD 742  
Db 673 FEERADKDSATYAGALRPYRARDLQVDELQVETEIDWLGDG-----LLELTSHISYD 725  
QY 743 NWGQ-----IANTHSYGVSEKITVDPIITL-----ATKLOSNNNVQTGKVTTPSQ 793  
Db 726 DWGRYAVLPDGTQVEE---IDQVASTDGPVHRAWREVE---HSRVSGITETWNLFE 779  
QY 794 OPTQITLDFEAGHLQ-SCHTLTRDGDWRVEKETDAIG--OCTIYOYDNNVRIQITLDP 849  
Db 780 KPVRIERFALDGTSTISLQVNDYDGLRLSREAGGAGGQVVEYRIDAFDRVERLAD 839  
QY 850 G-TIVNRKYAPFSTDLITDIRVN-----GISLQOQTFDGLSKLTOSQDGRVWAYTSA 903  
Db 840 RVNTVRYTGHRSRNDLPVSIKVGNTESAAVLGEOVFDGLERRTVAITGGRQCTPEYDP 899  
QY 904 GNDQCPSTVTPDGOFTHYQYQBEL-DDAVLOVASNEITQOFSYNPVPGALLKAVAGQS 962  
Db 900 GERQ-PHWKAPDGTRIEYOYRPAIGEEPVLRLVLSGK-EAKYBYDLKNAKLTHCEBPDD 957  
QY 963 LTPY-----YPSGRLKMN-----INDMKMSYLMVTLRLENGYDLDLTIQIKSRDTH 1012  
Db 958 ENSGYTLDRSHFLSNGEVEKESRTVDGEAFSMYDYFSFRSLRAYVDVLQOTLYDFDDV 1017  
QY 1013 GRVTOIKDSS-----IKTLNYYDLNRRHIGSQVDTLACHMLTTT 1052  
Db 1018 GRLEKTTLHAPEKPRTRYLRQAPARQLLLESTFGYDQGRVASITTTDASTGHATL 1077  
QY 1053 VEPDLNREIGRLKCDSSGHTLDIQOSWLKTQOLANRIVK-----LNGVLQRTQOYS 1104  
Db 1078 LEYDEFREILRF-DFGDTVQTLAQDYDFEDCLKSLRILKRPKGSDESQATLLRHETQ 1136  
QY 1105 YDSRNLNRYKCDGAECPDKYGHSHIVTQNFYDIYGNITACHTTFADGTEDHATPKFAN 1164  
Db 1137 YDRGRQLIVTCDGPEAPVDPGQTIARQIFGFDGLNLSLVITYRDPGSGWQRTLYEFKN 1196  
QY 1165 PTDPQ-----LTVVHTHPDMPD-----NRLKYDKAGRVINITHGN 1204  
Db 1197 -SDPAQMSRIIPDALLTETVDHDLBELPKVVKYDPLQALIDLHYDNGNL---ISDEQGR 1253  
QY 1205 TENFYDTLGRL-----QNGQSVYGYDPLNRL 1232  
Db 1254 V--LTYDGLNRLRVEPDPGERCRNYDYPENIL 1284  
RESULT 8  
Q883W6 ID Q883W6 PRELIMINARY; PRT; 982 AA.  
AC Q883W6;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE YD repeat protein.  
GN PSPTO2231.  
OS Pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
RN NCBI\_TaxID=323;  
RP [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,  
RA Dodson R., Deboy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkac L., Beanan R., Haft D., Selengut J., Nelson W., Davidson I.,  
RA White O., Fraser C., Collier A.;  
RA "Complete sequence of Pseudomonas syringae";  
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AE016863; AA055747.1; --  
DR TIGR: DSPTO2231; --  
DR InterPro: IPR006530; YD.  
DR Pfam: PF05593; RHS\_repeat; 6.  
KW Complete proteome.  
SQ SEQUENCE 982 AA; 106800 MW; 00E7CA6B93E19B78 CRC64;  
Query Match 10.5%; Score 932.5; DB 16; Length 982;  
Best Local Similarity 29.7%; Pred. No. 6.7e-42;  
Matches 285; Conservative 142; Mismatches 357; Indels 175; Gaps 33;  
QY 642 ANTLTYDYELNLODDNRPFPVITTDVNGQLNRPFDGAGRVHYSQCLKSDGCKFVTI 701  
Db 31 SSTLCYRD-----DWNQR--CCITTDN-VQTYEYSDPIGSDVHK-----GPIQKT 74  
QY 702 HTQYDQDEGRHHTSYDYLTNGRQOTDPDKVHLSMSKSYDNMGQIANTHSYGVSEKIT 761  
Db 75 WKQSGDPEGR-----ISGRSET-----WNLN-----FGKPDRI 103  
QY 762 VDPIITLTKLOSNNNVQTGKVTTPSQOITLFDGAGHLSCHTLTRDGDWRV 821  
Db 104 ---TLTAGTGRSRTHSMSRSRLTT---TEQEL-----SRQFLYDGLGRC 144  
QY 822 RKETDAIQCTIYQYDNNVRIQITLPGTIVNRKYAPFSTDLITDIRV---NGIS--- 875  
Db 145 TEQDALQOQSTLFSYDNNRSMVSSFLADGVSINRSYAPQSSSELATMLEVHQNTTIV 204  
QY 876 LGQCFDGLSRLTQSDQGRVWAYTSAQNDQCPSTVTPDGOFTHYQYQBELDAVLQV 935  
Db 205 AGTQKFDGLERVQTKTGDVREQVFNADGEMQ--PESRTTAGLDNINFTYTRALTQIFSS 263  
QY 936 ASNETIQFSYNPVPGALLKAVAGQSUTPIYPSGRLKMNINDKMSYLM-----T 989  
Db 264 TAPDETAKFDYDKTSARLIEATNPQGTTRYDVHNLQTLGETWDLN--LGQAWETRHOSS 321  
QY 990 LRGLENGYDGL-----TGTQIKSRDTHGRVTQIKDSSIKTTLNYYDLNRRHIGSQVDTLA 1044  
Db 322 LLGRPIKRTDCLKGEAAGAETRYDYTLGRIRFINQSNLRTIIDVGLCKVATEDLQ 381  
QY 1045 TGHMLTTTVEFDGLNREIGRLKCDSSGHTLDIQOSWLKTQOLANRIVKLVGLQRTQOYS 1104  
Db 382 AGTVIIDMEYDDQOEILRTQTASNAALTLTQWAVDGLLKLTRDLQOAGSPLLHETFS 441  
QY 1105 YDSRNLNRYKCDGAECPDKYGHSHIVTQNFYDIYGNITACHTTFADGTEDHATPKFAN 1164  
Db 442 YDPRGLRTLNVYLSGLSPRDELQREMTQIFSFDELDNITLCQRTFTDGTSAARAFKYS 501  
QY 1165 P-----TDPQCOLTEVHTHPDMPDNRLKYDKAGRVINITHGNNTENFTYDTLGRL--- 1216  
Db 502 PGDDKHKRCQQLSTAYTPPKTPPTPSYDANGN--QLKDEHGN--SLHYDSQSRLQV 557  
QY 1217 -QNGQG--SVGYDPLNRLVSQKTTLDCEL---YYRETMVNEVNGEMIRLLRTGETII 1271  
Db 558 AETGAPISQYRYDGHGQLVATR-DGNESEILRFYEGHQLSTQVQEDQRTQYLHGERQPL 616  
QY 1272 AQQ---RASKVLLTGTDSQSVILTSKQNLQSEAYSAYGKHKSSTANDASILGYNGERAD 1328  
Db 617 GQOIVDDAEQTLILLITANDQSVMGFFQOQLEKAVYSAYGERHSEALLSTAGFNGVRE 676  
QY 1329 PVSQVTHLNGVRSYDPTLMRPHTPDSLSPPFGAGGINPYSYCLGDFINRSDPSGH----- 1383

```

Db 677 AANGWYLLNGGYRAYNELLMEFHSFDFLSFPAEGGVNPNYTCNLPALRDPDTHGDASGQ 736
Qy 1384 -----LSWQ-----AMTGTGMAIA-----GLLLTIATGMAIAAAGI-----1416
Db 737 TGRLLRDEGALPMQGGGDMGHWGVGVFVLGVAAITATLTATVTPVTVLGI 796
Qy 1417 -----AAAIASSTTTALAFG-----ALSVTSITSIVSGALEDASPKASSILGWVS-----1462
Db 797 SMTASAAAAGVSTGALIVGTALTAASTTANTVAIVNN-----DOTAGEVGVGWLGAIV 851
Qy 1463 ---MGMCAAGLAESAIGKGTFLA-----THLGAFAEDGENALLKSTSSSR 1505
Db 852 PVLGVFGAGAVARAVAAAKVAANAGTIGVRSVRIG-LAAAGARTISSAASAR 909

RESULT 9
ID Q87VG6 PRELIMINARY; PRT; 1669 AA.
AC Q87VG6
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE YD repeat protein.
GN PSPT04970.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OK NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Beery K., Utterback T., Van Aken S., Feldblyum T., Gwin M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Hatt D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.,
RA "Complete sequence of Pseudomonas syringae."
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE016874; AAC583398.1; -.
DR TIGR: PSPT04970.
DR InterPro: IPR005530; YD.
DR Pfam: PF05593; RHS_repeat; 5.
DR Complete proteome.
SQ SEQUENCE 1669 AA; 184268 MW; BC668B6C1E3FB7EE CRC64;

Query Match 10.4%; Score 920.5; DB 16; Length 1669;
Best Local Similarity 25.9%; Pred. No. 7e-41;
Matches 440; Conservative 212; Mismatches 662; Indels 387; Gaps 84;

Qy 152 AFDLKVPKLLNPAGHAIYIDMFEATQPELNRIYDDLGDH:PLLNLEYQGL---IKTI 208
Db 72 AFGQLPSEVTSPOGGTSLSEWSRNOQOFLKKI--SSDGNVI--LSSTYSAPSSDTSTV 127
Qy 209 -LTLFPQKGYRTELRFANRLQSLNHNFSLGNENPLTWSFGYTPIGKNGILGOWITSM 267
Db 128 EISVWPKTDEAYTVRLSELSTALGTTRIEKNLKVSKLKYGC---ADPTLDRVLNRIE 184
Qy 268 APGLKELTVNYSNNQGHFPQSANLPVLPVYVTLMKQVPGAGAPAIQAEYSYTSNHYVGG 327
Db 185 EEDGSIELVYREG--GMPPFYRQ--PPLPCVTLHSLFPGACQSNITDHYFSGTNYLGF 241
Qy 328 GSGINGWKNLDNLYGLMTEYNYSSTSRRYKDEGH-----DQIVRIERTYNNYL 378
Db 242 SEPP--DAHONRLY-----YERLELRELVDDGYVQLRQNPDIHIVSTRHAFKNHL 292
Qy 379 LTSECKQONGYIQTETAYVIAIGHNFDSPQSQ--FQLPK--TKTETWSADNSYRSEIT 434
Db 293 QVRE-----DLQVQFOAEKVISWEFANASPGKVFGLPKITITDYLSHPNTERTTV 346
Qy 435 ETTFDESGNPLTKVIKDKTKQKILSPSTHWEYYPAG-----EVD-----NC 476
Db 347 QTLAYNNIGQLTKSI-----AVDGVVTEWLYYPTDGGQGLDISLIAEKPLSKDLVTLTC 400

```

```

Qy 477 PPEPYGTFRFVKKTIQTTPYDSEFKDDP-----EXFIOVRY-----SLIGS 516
Db 401 PKVSEGY-----MPVKVEYVHDPAYPKSKQITAVAYQARENPNVQSVLPSTVVV 452
Qy 517 QSHVTLKIEERHYSATQLLNSTLFOYNTDKSELGRLLKQTECTKG-----ENKTY--S 568
Db 453 LTGVTL-----DPTTMLPSLM-----EGKNALIEQRVISRIPDVAENTTAWKES 499
Qy 569 VVHKFTYTKQDDTLQCSHSITTHDNFTIHR-----SOVSRYTGRLESDTDTK 616
Db 500 VVQNSWLWCQNRNL--TTSMLYDDNPSVGRVRAEAGKIISSRIFSELSCRPLSET-RD 556
Qy 617 DIVTQMSYDKLGRLLTR---TLNSGTFYANTL-TYDIEL--NNLODNNRPPFVITTDVN 670
Db 557 GLEFHYVHDSLGRIVRQERGTEAGAWKADAVETTDYSITAEGLO-----VTVTEAE 608
Qy 671 GNQLNEFDGAGRHSVQCLKDS--GDGKFYTIHQOYDEQGRHHTSTVSDYLTNGROTD 729
Db 609 -QQVRLYDGLQRPVWVAIKTILPDSAPCVISIEYDGPDPATNQTLY-DYLPGLRRTK 666
Qy 730 PKVHLSMSKSYDNWGQIANTHWSYG--VSEKITVDPIITLPAKQLOSNNVOTGKEVT 787
Db 667 DARPEAAVDASKLAW-MADYTRERDAGILNEQVIGADSGAQLIRQLSGRNLDSHTALLE 725
Qy 788 TYTPS-----QOPIQITLFDAGHLQSCHTLTRDGDWRVRKETDAIGCTIYQYDNYN 840
Db 726 TLRPSTARDASTDRTIERT-FDERRLIKRT-----SNTSEHC--JEYDELE 770
Qy 841 RVIQITLPGDTIVNRKYAPFSTDLITDIRVNGISLQQOTPDGSLRSLTQSDQGRVWAYT 900
Db 771 RAVALLIADPTGRTERKXHOLS--DYITQLNVGTVLGTQKMTAAARQTTVGE-----LTYE 824
Qy 901 YSAGNDQCFSTVITPDGOFIHYQVQPELDDAVLOVASNEITQQ-----FSYNPVT- 950
Db 825 PFGGS---ASTVVRPDKTLESASVADGHTATLSINKKVHTQRVISQPNVLTVTVDPVS 881
Qy 951 ----GALLKAVAGCSLTPYY--PSGRKLMENINDMKMSYLTLRLGLENGYTDLTGT 1003
Db 882 PSAEAWSSLTSSPOSGLGTSITQTSPRGSRQAEWTRSLKG-----RLLTN--TAVDGR 932
Qy 1004 IQKISRDRTHGRVTOIKSSIKTLYNDLNRHIGS-----QVTDLATGHMLTTTVEFGLN 1059
Db 933 QMRVFRDYLDVRVVRV---ILGELHYLHLSAFGEPLQRTVNVQASGERLDVFTWDAPG 988
Qy 1060 REIGRKLCDSGHTLDIQOSWLKTQQLANRIVKLNGVLQRTQOYSYDSRNLNOKYC--D 1117
Db 989 QEIAREYTLNNKPLALNTNSYLANGQVSSKTLTREGVLQRTGFSYDARDLNSYECTTD 1048
Qy 1118 GAEPTDKYGHISIVTQFTYDIYGNITACHTTFADGTHEDHATKFX---ANPTDPCQL-TE 1173
Db 1049 VADWPQDQAGSKLSQSYGYDELHNLSECSSTYADGSTCIQTYTYTVTKNPTRLSVKTE 1108
Qy 1174 VHH--THPMDPNIRLKYDKAGRVINITDNHGNTEFTYDILGRL---ONGQGSV---YGY 1226
Db 1109 LRSSQTTSTQATLADANGN--QTTDESQRT--LAYTPLGLASVKDNDKLLITYSY 1164
Qy 1227 DPLNRLVSQKDTTL--DCELYRETMLVNEV-----RNGEMIRLLRT-----GETII 1271
Db 1165 DAFGRLLSQYIGATKHTCELLYDGTQLTGEAWFDANREFKRILFSDVMVQTCIGETV- 1223
Qy 1272 AQQRASKVLLTGTDSQOSVILLTDKQ-----NLSQEAYSAGKHKSTANDASI-LGVNGE 1325
Db 1224 ---RSDFVL---TDPGGVGVFSADHGTAGVKLHPLGYTPYG--ESTNLSGGRLGNSE 1275
Qy 1326 RADPVSGVTHLNGYRSYDPTLMRFHTPDSLSPFAGGINPYCYCLGDPINRSDPPSGLH- 1384
Db 1276 RIDPVLGWHLNGYRTYSPAQRHQLQDPSWSPFAGGINNTAYCAGDPVNLFPDPSGHVM 1335
Qy 1385 --SQA-----WTGI-----CMGIAGLLITATGMAIAAAGGI 1416
Db 1336 ISRWGASNMISDLTKALQETSPOQLHFWRGLAVNASVAVAGVLMVPLTGGSSLGPAAGV 1395

```



```
QY 488 KKIQTPTVDSF-----KDDPKFTQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQYN 543
Db 455 KQIVTPADSEFYAPVQQTYYAYAQYPCIAAGSSLSYAVLQTOETLCSDDVL-----LLTIN 510

QY 544 TD 545
Db 511 TD 512

RESULT 12
Q52883 PRELIMINARY; PRT; 526 AA.
AC 052883;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN ORF 526.
OS Coxiella burnetii.
OG Plasmid QpRS.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Priscilla 0177;
RA Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y15898; CAA75844.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 526 AA; 60104 MW; 42D04FF8CC361CC CRC64;

Query Match 10.3%; Score 916.5; DB 2; Length 526;
Best Local Similarity 40.8%; Pred. No. 1.8e-41;
Matches 221; Conservative 73; Mismatches 203; Indels 45; Gaps 16;

QY 18 NEF-FTQANNFTSAVSGVDPRGTGLYNIQTILGHVGNGLGPTLPLTSLYSPLNKTDI- 75
Db 2 NELPQTQATNFISAVQGVDPRTGLFTVNMVLAELTGNDNLGDPDLFTLNSPLSTNIC 61

QY 76 GFGIGFNLGVYDRKNSLSLSTGKNTYKTIETDKTVKLOOKLNDLRFKDLKNCYRI 135
Db 62 GFGIGCSVGISIYDKNNKLLLSGGERYKTEDMDNGVYVROKINNPFKEK--IKNGYII 119

QY 136 IHKSGDIEVLGTG--FNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDLDGHDIP 195
Db 120 KYKNGKTEYLYKYGDNLF---LPQKIFSTLWPLKLTWENRGQVYNLKIEDAKD---V 172

QY 196 LLNLEYQGLIKTILTLFPQOKGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTPIGK 255
Db 173 LCKIDYQPSDWARITFWPGKTESYTFQDLFVNELYYVWTKNKSRE--LWVSFNDDVGA 230

QY 256 NGILGQWITSTAPGGLKETVYNNNGHFFQSANLPVLYVTLMKQVPGAGQAIQA 315
Db 231 GNFT---LTQVKSSTGLTETVNYQAGV--RFPDESGKPALPSVNYRQSPGQPDIVK 285

QY 316 EYSTSYNNYVGGGN--GIWNNKLDNLYG-LMTEYNTGSTESRYKDKGHDQIVRIERT 372
Db 286 EYEYTAGNLYGASLGAKWNEDEDNINVMDDYTSSTEKLIVDNRE---LVSIISRI 341

QY 373 YNNYHLLTSECKQNGYIQTETAYVAIIGHNFDSPQSPQFOLPKTKTETWR--SADNSYRS 431
Db 342 YNSYLLISETRQNSCEVIVETDYAKPGLSPDKQKQQLPKKEKKTWRENSKNQCRS 401

QY 432 EITTTFDESNGPLTKVKKTKQIKIISPTSHWEYYPAGEVDN-----CPPEYGFTRFV 487
Db 402 EITTTTFDPEGNLLTKIEPD-----GKTEYIYDYSKGETDKGIVLCPPPEPNGFVRV 454

QY 488 KKIQTPTVDSF-----KDDPKFTQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQYN 543
Db 455 KQIVTPADSEFYAPVQQTYYAYAQYPCIAAGSSLSYAVLQTOETLCSDDVL-----LLTIN 510
```

```
QY 544 TD 545
Db 511 TD 512

RESULT 13
Q45949 PRELIMINARY; PRT; 528 AA.
AC 045949;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Orf 528.
OS Coxiella burnetii.
OG Plasmid QpH1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I;
RA Thiele D., Willems H., Haas M., Krauss H.;
RT 'Analysis of the entire nucleotide sequence of the cryptic plasmid QpH1 from Coxiella burnetii.';
RL Eur. J. Epidemiol. 10:413-420(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I;
RA Thiele D.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; X75356; CAA53130.1; -.
DR PIR; S38242; S38242.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
KW Plasmid.
SQ SEQUENCE 528 AA; 60422 MW; 492589F2DAA890A6 CRC64;

Query Match 10.3%; Score 912.5; DB 2; Length 528;
Best Local Similarity 40.8%; Pred. No. 3e-41;
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

QY 18 NEF-FTQANNFTSAVSGVDPRGTGLYNIQTILGHVGNGLGPTLPLTSLYSPLNKTDI- 75
Db 2 NELPQTQATNFISAVQGVDPRTGLFTVNMVLAELTGNDNLGDPDLFTLNSPLSTNIC 61

QY 76 GFGIGFNLGVYDRKNSLSLSTGKNTYKTIETDKTVKLOOKLNDLRFKDLKNCYRI 135
Db 62 GFGIGCSVGISIYDKNNKLLLSGGERYKTEDMDNSVYVROKINNPFKEK--IKNGYII 119

QY 136 IHKSGDIEVLGTG--FNNAFDLKVPKLLNPAGHAIYIDWNFEATQPLNRIYDLDGHD 193
Db 120 KYKNGKTEYLYKYGDNLF---LPQKIFSLGFWPLKLTWENRGQVYNLKIEDAKD--- 173

QY 194 IPLANLEYQGLIKTILTLFPQOKGYRTELRFNLNQLNSIHNFSLGNENPLTWSFGYTP 253
Db 174 -VLCKIDYQPSDWARITFWPGKTESYTFQDLFVNELYYVWTKNKSRE--LWVSFNDDV 230

QY 254 GKNGILGQWITSTAPGGLKETVYNNNGHFFQSANLPVLYVTLMKQVPGAGQAI 313
Db 231 GAGNTT---LTQVKSSTGLTETVNYQAGV--RFPDESGKPALPSVNYRQSPGQPD 285

QY 314 QAEYSYTSNNYVGGGN--GIWNNKLDNLYG-LMTEYNTGSTESRYKDKGHDQIVRIE 370
Db 286 VKEYEYTSNLYGASLGAKWNEDEDNINVMDDYTSSTEKLIVDNRE---LVSIIS 341

QY 371 RTYNNYHLLTSECKQNGYIQTETAYVAIIGHNFDSPQSPQFOLPKTKTETWR--SADNSY 429
Db 342 RIYNSYLLISETRQNSCEVIVETDYAKPGLSPDKQKQQLPKKEKKTWRENSKNQ 401

QY 430 RSEITTTTFDESNGPLTKVKKTKQIKIISPTSHWEYYPAGEVDN-----CPPEYGFTR 485
Db 402 RSEITTTTFDPEGNLLTKIEPD-----GKTEYIYDYSKGETDKGIVLCPPPEPNGFVR 454

QY 486 FVKKIIQTPTVDSF-----KDDPKFTQYRSLIGSQSHVTLKIBERHYSATOLLNSTLFPQ 541
```

```

455 FVKTIQVTPADSEFYAPVQOTTYYAQAQPCIAAGSSLSYAVLQTOETLCSDDVL-----LIT 510
452 YNTD 545
511 INTD 514

RESULT 14
Q45902 PRELIMINARY; PRT; 589 AA.
AC Q45902;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Orf 589 protein.
GN ORF 589
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Scury 0217;
RX MEDLINE=97294470; PubMed=9150226;
RA Williams H., Ritter M., Jager C., Thiele D.;
RT "Plasmid-homologous sequences in the chromosome of plasmidless
RL Coxiella burnetii scury 0217."
RL J. Bacteriol. 179:3293-3297(1997).
DR EMBL; X93204; CAA63679.1; -
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs repeat; 2.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
SQ SEQUENCE 589 AA; 64241 MW; 8B6BEE8BEDE9A035 CRC64;

Query Match 10.0%; Score 884; DB 2; Length 589;
Best Local Similarity 38.0%; Pred. No. 1.3e-39;
Matches 232; Conservative 85; Mismatches 182; Indels 112; Gaps 24;

QY 1103 YSDSRNRLNQYKCDGACETDKYGSIVTQNTFYDIYGNITACHITTFADGTEDHATPKF 1162
DB 2 FAYDSRNLLEYTCNGEARPQDPYGAIRHQTFSYDALGNMTKTQTDPSGG-RNTATYIY 60

QY 1163 ANPTDPCQTEVHHTHDMDPNDRLKYDKAGRVINITDNGHNTENFYDTLGLQ--NGQ 1220
DB 61 -SAIDPTQLKVNNDHNYKEITLVDYDKAGMR--DEAGRT--LRYDALGLQVNGE 115

QY 1221 GS---VYGYDPLNRLVSKQT-DTLDCELYRETMLVNEVR--NGEMIRLLRTGETIIAQ- 1273
DB 116 GAKGGQAYDALNTLVSVQVQDEPIYDLYRADDLVGEARRDGSSTRYVKSQCCVGC 175

QY 1274 --ORASKV-LITGTDSCQSVLTSDKONLSCEA--YSAYGKHSTANDASILGYNGERADP 1329
DB 176 TKQSSNTSWLTTNQQSVLSVSEGNHAFQDCIYTPGYRTQTETPSVLGNGERLDP 235

QY 1330 VSGVTHLNGYRSYDPTLMRFHTPDSLPFGAGGINPYSCLGDPINRSDPSGHLGNQAW 1389
DB 236 VSSYVHLNGYRAYNPILMRNCPDSWSPFGAGGINPYAYCDGDPINRVDPNHLSQAE 295

QY 1390 TGIGMGAGLLLTATGMAAAGGTAAGTAIASTTALAFGALSYSDDITSTVSGALED 1449
DB 296 LGIGLVGVGLVAVFTAGTSIAAAGASAAIESASISLVVGTIGVAADVASIAGALE 355

QY 1450 ASPKASSTLGMVSMGMGAAGLAE---SAIKGKTGLATHLGAFAEDG-----ENALL 1497
DB 356 ANPOASATLWISLGLGPGGAVSLATAARAGKL---ISLAKGGGKIRSQSPVQGINY 412

QY 1498 KSTSESSIRKMGVT--RSLDREIVRNEE-----GOVTKHSRGYTNF----- 1538
DB 413 RSLSRGDPRLGGPPHFGSLSRTVAPESMRPAGLNHYKVSQKSLGYQHVFGADREIFG 472

QY 1539 -----MGKEQAII-----VHGDKDGFVHTEG-----NK 1563

Db 473 YEIREPIEFFRRRSITKRDIVILSGTHGRVHGDN-----WTSQGLRRPDILERAIFYEDV 528
QY 1564 HNGXGPYTRHTPEQLVDLYKDNNDIVLDTQGG-----DKPVHLLSCYKSSGAADKMA 1615
Db 529 QNYXG-----QLNGRVKIVDMAGMTESEFGRHVRNVNQHVILGYCWRNDEA---LT 577
QY 1616 KYIN-RPVIAI 1625
Db 578 YHRLRPIISY 588

RESULT 15
Q931J8 PRELIMINARY; PRT; 762 AA.
AC Q931J8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Vibrio sp. CH-291.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=161725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CH-291;
RA Jones J., Lewin A., Appel B.;
RT "Cloning of a hemolysin encoding region of a Vibrio species.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ314791; CAC40978.1; -
DR InterPro; IPR006530; YD.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW Hypothetical protein.
FT NON_TER 762
SQ SEQUENCE 762 AA; 87823 MW; A172ABBCD447D2CA CRC64;

Query Match 7.3%; Score 648; DB 2; Length 762;
Best Local Similarity 26.1%; Pred. No. 1.2e-26;
Matches 222; Conservative 150; Mismatches 341; Indels 136; Gaps 32;

QY 22 TOANFTSAVSGVDPRGLYNIQITLGHIVGNGNLGPTLPLTISYSLNKTIDIFGIGF 81
DB 6 SNAFNFSFVSGVDPRGTSYISITLGLLSNKLKSGNFKLITISHYLNKVDGEGFLGW 65

QY 82 NPLGSVYDRKNSLLSLSTGENYKVI---ETDKTVKLOQKLDNLF-----EKDLKENCVR 134
DB 66 SISMSSYDKISKRLSLSSGRTPETVLSDKSNELIILHRKTKDRAFLVENEREIK- 120

QY 135 ITHKSGDIEVLTFGNNAFDLKVPK--KLLAPAGHAIYIDWNFEATQPLRNKIVDDLDGH 192
DB 121 VYVIDGKVEYI-----DYESGKLIETVSNLGHIEIFYRYFNGLLSLSLSDQF-GH 171

QY 193 DIPL--LNLEYQGLIKTILT-----LFPQCKEGYRTELRLNRLNLSIHNFSLGNE 242
DB 172 SITIDHNSKYYTIINSVAENDIYKRNVLKLSHGY-----GRILTSI---SWANSS 221

QY 243 PLTWFGYTPIGKNGILQW-ITSMATAPGLKETVYNNQCHHFP-QSANLPVLPLYVT 300
DB 222 NLKTTIDYKIDK---LGNYAIQVKHYSGLVETIEYS--YEGHLLPKKNNFNPIPNVK 276

QY 301 LMKQVPGAGQPAIQAEYSYTSNHYVGGSGNWNKLDNLYGLMTEYNYGSPESRRYKDX 360
DB 277 RHILTPGSGQPKTVFEYYSLSKNLYGYSNLLWEGVDVLFYTADYYSQEVN----N 332

QY 361 EGHQDQIVRIERTYNNYHLLTSECKQNGYI-OTTATYAIIGHNFDQPSQFQPLPKYT 419
DB 333 GSH-----IIRVYVNVKHLLESEKFRNGVLYKEIDLEYVADLTQCIDQPNNYSYKKES 388

QY 420 ETWRSADNSYRSEIETTFDESGNPLTKVTKDKTKQKLIISPSTHWEYYPAGEVNCNPE 479
DB 389 ITHF-LDGTQRTSEVSEYGFDDYGCNCFECDDGTT-----SFFEYPCESQEQCPAH 440

QY 480 PYGFTFRVKKIITQPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIBERHYSATQLNSTL 539
```

[illegible]

Search completed: July 3, 2004, 05:50:11  
Job time : 77 secs